DOCUMENT RESUME

ED 041 751 SE 009 203

AUTHOR	Dash, J. G .
TITLE	Heat Motion in Matter.
INSTITUTION	Commission on Coll. Physics, College Park, Md.
SPONS AGENCY	National Science Foundation, Washington, D.C.
PUB DATE	65
NOTE	39p.; Monograph written for the Conference on the New Instructional Materials in Physics (University of Washington, Seattle, 1965)
EDRS PRICE	EDRS Price MF-$0.25 HC-$2.05
DESCRIPTORS	*College Science, Heat, *Instructional Materials, *Kinetic Molecular Theory, Motion, *Physics, *Probability Theory, Resource Materials, Statistics

ABSTRACT

This monograph was written for the Conference on the New Instructional Materials in Physics, held at the University of Washington in summer, 1965. It is designed for college students who are non-physics majors, and is organized in sections of increasing sophistication. Section 1 presents ideas related to the kinetic theory of gases. Section 2 deals with games of chance, precision measurements, quality control, and biological variation. In section 3, the binomial and other distributions are discussed. The statistics of pure chance are translated into the statistics of a gas system in section 4. The distribution in space of gas molecules, molecular velocities, phase space, distribution in phase space and the Maxwell-Boltzmann gas are topics treated in this section. The monograph concludes with a bibliography of suggested readings. (LC)

FILE COPY

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

Heat Motion in Matter

COMMISSION ON COLLEGE PHYSICS
DEPARTMENT OF PHYSICS AND ASTRONOMY
UNIVERSITY OF MARYLAND
4321 HARTWICK ROAD
COLLEGE PARK, MD. 20740

ED041751

J. G. DASH

University of Washington

SE 009 203

GENERAL PREFACE

This monograph was written for the Conference on the New Instructional Materials in Physics, held at the University of Washington in the summer of 1965. The general purpose of the conference was to create effective ways of presenting physics to college students who are not preparing to become professional physicists. Such an audience might include prospective secondary school physics teachers, prospective practitioners of other sciences, and those who wish to learn physics as one component of a liberal education.

At the Conference some 40 physicists and 12 filmmakers and designers worked for periods ranging from four to nine weeks. The central task, certainly the one in which most physicists participated, was the writing of monographs.

Although there was no consensus on a single approach, many writers felt that their presentations ought to put more than the customary emphasis on physical insight and synthesis. Moreover, the treatment was to be "multi-level" --- that is, each monograph would consist of several sections arranged in increasing order of sophistication. Such papers, it was hoped, could be readily introduced into existing courses or provide the basis for new kinds of courses.

Monographs were written in four content areas: Forces and Fields, Quantum Mechanics, Thermal and Statistical Physics, and the Structure and Properties of Matter. Topic selections and general outlines were only loosely coordinated within each area in order to leave authors free to invent new approaches. In point of fact, however, a number of monographs do relate to others in complementary ways, a result of their authors' close, informal interaction.

Because of stringent time limitations, few of the monographs have been completed, and none has been extensively rewritten. Indeed, most writers feel that they are barely more than clean first drafts. Yet, because of the highly experimental nature of the undertaking, it is essential that these manuscripts be made available for careful review

by other physicists and for trial use with students. Much effort, therefore, has gone into publishing them in a readable format intended to facilitate serious consideration.

So many people have contributed to the project that complete acknowledgement is not possible. The National Science Foundation supported the Conference. The staff of the Commission on College Physics, led by E. Leonard Jossem, and that of the University of Washington physics department, led by Ronald Geballe and Ernest M. Henley, carried the heavy burden of organization. Walter C. Michels, Lyman G. Parratt, and George M. Volkoff read and criticized manuscripts at a critical stage in the writing. Judith Bregman, Edward Gerjuoy, Ernest M. Henley, and Lawrence Wilets read manuscripts editorially. Martha Ellis and Margery Lang did the technical editing; Ann Widditsch supervised the initial typing and assembled the final drafts. James Grunbaum designed the format and, assisted in Seattle by Roselyn Pape, directed the art preparation. Richard A. Mould has helped in all phases of readying manuscripts for the printer. Finally, and crucially, Jay F. Wilson, of the D. Van Nostrand Company, served as Managing Editor. For the hard work and steadfast support of all these persons and many others, I am deeply grateful.

Edward D. Lambe
Chairman, Panel on the
New Instructional Materials
Commission on College Physics

H E A T M O T I O N I N M A T T E R

J. G. Dash

C O N T E N T S

1 ELEMENTARY KINETIC THEORY OF GASES 1

2 CHANCE, ERROR, AND STATISTICS 6

2.1 Games of chance 6
2.2 Precision measurements 8
2.3 Quality control 9
2.4 Biological variation 11

3 AN INTRODUCTION TO THE MATHEMATICS OF PROBABILITY 13

3.1 The binomial distribution 13
3.2 Other distributions 18
3.2.1 The skew binomial distribution: a "plugged nickel" 18
3.2.2 The multinomial distribution 19
3.2.3 The Poisson distribution 19
3.2.4 The Gaussian distribution 19

4 POSITIONS AND VELOCITIES OF MOLECULES OF A GAS 21

4.1 The distribution in space 21
4.2 Molecular velocities 24
4.3 Phase space 24
4.4 Distribution in phase space 26
4.5 The Maxwell-Boltzmann gas 28

EPILOGUE 33

BIBLIOGRAPHY 34

1 HEAT MOTION IN MATTER

One hears that seemingly stationary matter is really composed of atoms in violent motion. Somehow, the motion is caused by heat. It is often called "heat motion," but the name simply begs the question: Is "heat" just another word for motion? As the temperature is reduced the heat motion becomes less violent. Is heat different from temperature? At absolute zero the atoms are . . . motionless? The last question is usually especially puzzling to the student, who may feel that "absolute zero" has a forbidding sound. Some have heard that absolute zero is unattainable, which is forbidding enough.

How do we know that atoms have heat motion, and how does the motion affect what we can feel and measure? What happens near absolute zero? The answers are contained in the branches of physics known as kinetic theory and statistical mechanics. In this monograph the subjects are introduced by studying heat motion in a gas.

We begin with atoms, as Leucippus and his student Democritus (460-370 B.C.), assumed them to be, indivisible and unchanging. The atomic hypothesis, as it was formed into a philosophical system by Epicurus (341-270 B.C.), was the inspiration for the poem De Rerum Natura of Lucretius (died ca. 55 B.C.), who viewed all of inanimate nature, life, and society as increasingly complex systems progressively developed from the natural laws governing atoms. This poem has been called a versified textbook in atomic physics, and it presents a compelling unified view of nature. But Aristotle (383-322 B.C.), who had the greatest influence on the natural philosophy of later Greeks, found the atomic basis of the universe and our perceptions of it distasteful and incompatible with his own conviction of divine purpose. His criticism of the atomists seemed sufficient reason to the medieval philosophers for rejecting their theories, and Lucretius' poem was banned. By the time it was made widely known again, by the Commentaries of Gassendi (1592-1655), two thousand years had passed since Democritus.

Publication of the Commentaries revived the ancient discussions on the nature of air, which had played an important role in the Greek atomists' speculations on the physical influence of unseen things. Gassendi himself helped to set the stage for the modern study of gases. His atoms move freely in all directions, accounting for the free diffusion of the gaseous state in apparent violation of gravity. The particles have mass, and particles of like matter have like mass. Thus, the density of a gas at a particular pressure depends only on the average number in a unit volume. Being free to move in any direction, they collide with all walls and partitions. If a partition is moved in, so as to decrease the space available to their movements, the rate of atomic collisions with the partition increases, and therefore more force will be necessary to oppose their impact. With this qualitative model, Gassendi prepared the atomic hypothesis for the quantitative researches of Boyle (1627-1691), and Hooke (1635-1703), on the elasticity of gases. With this cursory view of its historical origins, we will begin our study at this point, in a simplified treatment of the gas laws.

Boyle (or Hooke), discovered, in 1660, that the pressure of a gas in a container varies inversely with its volume, so that the product PV for any given body of gas is a constant as the volume is changed. (It has since been named Boyle's law.) Daniel Bernoulli gave a quantitative explanation, in terms of atoms, in 1738. A simple derivation is as follows: We imagine a cubical container filled with a large

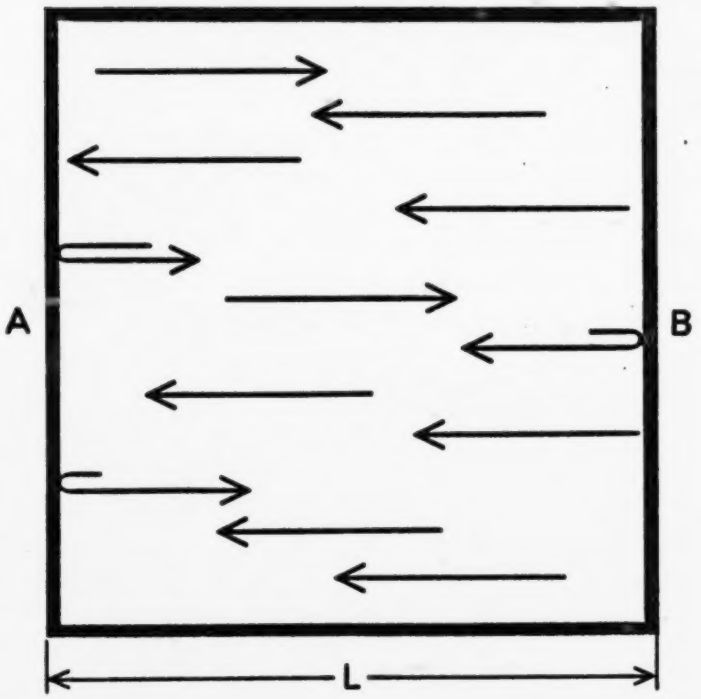


Fig. 1.1 Diagram of an oversimplified model of molecular motions in a cubical volume.

number of identical atoms. At first we imagine that all the particles move with the same speed v back and forth in each direction, as illustrated in Fig. 1.1. Each atom travels back and forth between faces A and B, its momentum changing by 2mv at each collision (its mass is m); the time between two successive impacts on each face being 2L/v. Hence, the impulse, or change of momentum experienced by all N particles at each wall is

$$\frac{\text{change in momentum}}{\text{elapsed time}} = N\,\frac{2mv}{2L/v} = N\,\frac{mv^2}{L} \qquad (1.1)$$

in unit time. This rate of change of momentum is accomplished by the application of forces at each collision, and clearly it is the walls that supply the forces: According to the Newtonian principle of equal action and reaction, the forces on the particles and on the wall are equal in magnitude, and opposite in direction. Thus, as the particles are turned back into the container at each collision, the wall is pushed outward. The total reaction on the wall is the total effect of a steady rain of impacts. Under most conditions, all that can be detected is a time-averaged force. The average force is equal to the average rate of change of momentum "delivered" to the atoms by the wall: It is just the value in Eq. (1.1) above. Because the force is imparted by many impacts it also seems to be distributed uniformly overthe surfaces of faces A and B. A unit area of each face experiences an average pressure P equal to the force per unit area of the face:

$$P = \frac{Nmv^2}{L^3} = \frac{Nmv^2}{V}, \qquad (1.2)$$

where $V = L^3$ is the volume of the container.

By using an extremely artificial model we have obtained the property PV = constant, for faces A and B. But what happens at the other faces? Our next step is to improve the theory by using a more realistic model. Instead of assuming that all atoms move back and forth between faces A and B, we should not expect that these two faces will be singled out, but that all six faces are equivalent, all experiencing the same average pressure. One way of establishing the equivalence would be by assuming that one third of all the atoms move back and forth between A and B, one third between top and bottom, and the one third between the remaining pair. In this way, all faces experience the average pressure

$$P = \frac{1}{3}\,\frac{Nmv^2}{V}.$$

Again, we can see that the model is unrealistic and can be improved. We should expect that atoms will move not only along the three directions parallel to the sides, but will also take all the intermediate directions, with arbitrary velocity components v_x, v_y, v_z, in a Cartesian coordinate system. The axes of the coordinate system can be oriented in any manner relative to the sides of the container. The total behavior of all the particles will

still be uniform in all directions: There is no preference for one direction over any other. We can imagine an "average atom" which traces out a complicated path which samples all directions in succession, so that its average velocity components have equal magnitudes:

$$|v_x| = |v_y| = |v_z|. \qquad (1.4)$$

Since the total velocity is the vector sum of the three Cartesian components,

$$v^2 = v_x^2 + v_y^2 + v_z^2$$

$$= 3\, v_x^2 .$$

The pressure on each face, due to the averaging over all directions of the atomic velocities, is the same as in Eq. (1.3). Thus, we see that the pressure is uniform on all faces, even if the atoms move in random directions, provided that all directions are equivalent.

Equation (1.3) can be rewritten in terms of the kinetic energy ϵ of an individual particle. Since $\epsilon = \frac{1}{2}mv^2$,

$$PV = \frac{2}{3} N\epsilon = \frac{2}{3} U_{kin} \qquad (1.5)$$

where U_{kin} is defined as the total kinetic energy of all the particles. Therefore, the product PV is constant for a constant total kinetic energy of the gas enclosed, as long as all particles have the same speed.

Clearly, this last assumption is unreasonable. We cannot accept the hypothesis that a gas of randomly moving atoms has only one characteristic speed. More likely, there are collisions between atoms which cause some to slow up and a few to gain most of the energy of the two; that is, there is probably some randomness in speeds as well as in directions. However, we might expect that there is some average[1] speed of the atoms that is unchanging. If the average molecule has an average speed then it has an average energy $\langle\epsilon\rangle = \frac{1}{2}m\langle v^2\rangle$, and Boyle's law is still satisfied:

$$PV = \frac{2}{3} N\langle\epsilon\rangle = \frac{2}{3} U_{kin}, \qquad (1.6)$$

where the total kinetic energy $U_{kin} = N\langle\epsilon\rangle$.

A simple extension of these arguments leads to another important result, known as Dalton's law. Suppose that the quantity of gas is composed of two types of atoms. The pressure at the walls is a result of the impacts of both varieties, which we take to have numbers N_1, N_2, and average energies $\langle\epsilon_1\rangle$, $\langle\epsilon_2\rangle$. Then, the product PV is, after Eq. (1.6), the result of both types of atoms bombarding the walls,

$$PV = \frac{2}{3} N_1\langle\epsilon_1\rangle + \frac{2}{3} N_2\langle\epsilon_2\rangle$$

which we can write in terms of two partial pressures P_1 and P_2, where P_1 and $P_2 = (2/3)N\langle\epsilon_1\rangle$, etc. If more varieties are present, the total pressure P is the sum of all partial pressures, where each partial pressure is equal to the pressure that each type would have if it alone occupied the container. Thus, for ν different atomic species, the total pressure is

$$P = \sum_{i=1}^{\nu} p_i, \text{ where } p_i = \frac{2}{3} N\langle\epsilon_i\rangle. \qquad (1.7)$$

Although we have derived the pressure equation for the special case of a cubical container, it can be easily extended to arbitrary shapes. For we can imagine any shape composed of a series of identical cubes, each filled with the same number of identical atoms having the same average speed $\langle v\rangle$. Every pair of walls of adjacent cubes would be subject to the same average pressure. Therefore, the internal walls are really not necessary, since they provide no net force, and there is no

[1] The symbol $\langle\ \rangle$ is used quite generally in physics to denote an average quantity. Somewhat less often, a bar over the quantity is alternatively used. For example, the average speed can be written $\langle v\rangle$ or $\bar{v}$.

discernible difference between two atoms rebounding from opposite sides of a wall and their passing through the wall without collision, with each one moving away from the wall into the other atom's container. Therefore, removing all inside walls will not change the pressure, since the total number of particles per volume N/V of the large composite volume is the same as for each of the component subvolumes.

Soon after the discovery of Boyle's law it became evident that it was only conditionally obeyed. Amontons observed in 1702 that the empirical constant increased if the quantity of gas was warmed. But the manner of its change could not be specified quantitatively without some manner of measuring how much warmer. Although crude thermometers had been invented by Galileo and others, no satisfactory instrument existed until Fahrenheit (1686-1736) developed a mercury thermometer, and established a temperature scale determined by the boiling points of liquids. With its aid, Charles (1746-1823) showed in 1787 that the constant of Boyle's law varies linearly with the temperature. That is, according to the temperature t on the temperature scale defined by a mercury thermometer, the variations of pressure and volume of a fixed quantity of gas can be described by the equation

$$pv \propto (1 + at), \qquad (1.8)$$

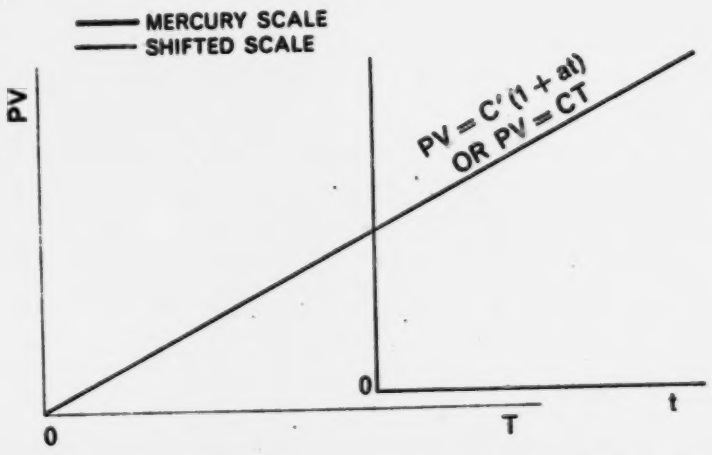


Fig. 1.2 Illustrating the relation between the temperature scales according to the mercury thermometer and the expansion of gases.

where C and a are constants. Charles' law can be used to establish a more convenient temperature scale, defined by the gas itself. If we define a new temperature T in terms of the mercury thermometer's scale by the relation

$$T \propto 1 + at,$$

then Charles' law becomes

$$PV = CT. \qquad (1.9)$$

where C is a constant. This simple transformation of the temperature amounts to a shift of the zero of temperature, illustrated in Fig. 1.2. As it was elaborated by Gay-Lussac (1778-1850), and others, the modern form of the gas law is given in terms of the number of atoms N and a universal constant k, the Boltzmann constant:

$$PV = NkT. \qquad (1.10)$$

The Boltzmann constant $k = 1.380 \times 10^{-16}$ erg/atom °K is related to the "gas constant" R by Avogadro's number $N_A = 6.025 \times 10^{23}$ atoms/mole and the number of "moles" n_m:

$$Nk = n_m N_A k = n_m R;$$

$$R = 8.317 \times 10^7 \text{ erg/mole °K} \qquad (1.11)$$

k and R are experimentally determined, and have been found to be universal constants, independent of the chemical species or mass of the gas. As long as the gas is not too dense or cold, the gas law provides an accurate description of molecular gases as well. It can be seen that an experimental law such as Eq. (1.10) might be used as the basis for a universal temperature scale, since it does not depend on particular details of the experimental apparatus. It is actually used for the establishment of an absolute temperature scale. The experimental and theoretical details of its establishment, although of fundamental importance in thermodynamics and statistical mechanics, is tangential to the main purpose

of this monograph. Interested students may read an extensive and lucid description in Heat and Thermodynamics, by M. W. Zemansky (see bibliography).

The gas law, established by over two centuries of experiments, put into its modern form, Eq. (1.10) leads us to a microscopic interpretation of temperature. In Eq. (1.6), we obtained an atomic basis for Boyle's law, in terms of the average kinetic energy $\langle \epsilon \rangle$ of the particles. It now seems, that if the product PV is found to vary with the absolute temperature, it must be that the average energy of the atoms is changed accordingly: comparing Eq. (1.6) and Eq. (1.10), we have

$$\langle \epsilon \rangle = \frac{3}{2} kT. \qquad (1.12)$$

Thus, with a combination of experimental fact and theoretical argument based on the atomic hypothesis, we have come to a beautifully simple and profound insight into the heat motion of the atoms of a gas. However, at this stage in our development, we cannot say that Eq. (1.12) is "proved," since it is not a purely mathematical theory. Its ultimate test can only be done by actually measuring the average energy of the atoms, and finding that they obey the equation. That test will in fact be the climax of our story. But before that, we will improve our theoretical model of the heat motions, to develop a distribution law for the velocities. This will lead us away from our direct concern for a while, to a description of statistical distributions in many other systems, and a short introduction to mathematical probability theory. We will return, in Chapter 4, to the microscopic model of a gas, and apply statistical methods to its analysis.

2 CHANCE, ERROR AND STATISTICS

There is a common thread in the mathematics, called statistics, which we use to study such widely diverse fields as gambling, quantitative measurements, biological variation, and the physical nature of matter. Although we want primarily to understand its part in physics at this moment, the roles that statistics plays in the other areas are also fascinating. Moreover, by studying its appearance in many guises we can hope to understand better how it works in one, and as we see it in each additional situation it shows us other aspects, becomes more fully fleshed, and familiar. In this chapter, we only talk about the statistical laws but do not use mathematics; in the next, the mathematical forms are derived and analyzed.

2.1 GAMES OF CHANCE

Successful gamblers know at least the simplest facts about statistics in their bones. They know that the laws of large numbers include not only the most elementary, that averages of groups of many identical things will agree, more or less, but also that deviations are bound to occur. Improbable runs of luck are possible; they are what makes the game interesting. Furthermore, improbable runs are bound to happen, if one plays long enough, and the frequencies of their appearance can be estimated. The expectation of probable and improbable runs to be found in the game can be determined from the distribution function of its results. The distribution function is simply a complete record of a very large number of sample games, so that all of the improbable events have had a chance to appear, and all of the more likely ones to appear many times, to establish a well-defined average behavior. If we have studied one group of many games, we know how future groups tend to work out; the distribution function is a property of the game, not the particular set of contests on which we base the statistical analysis. If the sample set contains few entries, it may be unrepresentative, containing an unusual proportion of unlikely results; To be characteristic of the game itself, the sample set should contain an infinite number of contests. Whether one tries to achieve very large sets in practice or not, the concept of probability involves the idealization that the infinite set has been played, or could be played; and the a priori probability of a particular result is the fraction of times that this result would appear. We say that the probability of a well-balanced and well-tossed coin coming up heads is 1/2; in a sense, that probability implies a definition for "well-balanced and well-tossed." We don't know that it is so until an experiment is done, but until it is, we accept the a priori probabilities themselves as representing the infinite set of trials.

It is amusing to read of the historic experiments designed to check the calculations based on a priori probabilities. The naturalist Buffon, assisted by a child tossing a coin in the air, played 2,048 partis of the Petersburg game, in which a coin is thrown successively until a parti is brought to an end by the appearance of heads. A Swiss astronomer studied the behavior of dice in several long sequences, which he analyzed for all of their distinguishable combinations; he wrote that altogether, in the course of his life, he had made 280,000 casts of individual dice. After he had completed the major part of his investigations, he found that the results were very different from the predictions of theory; some combinations were significantly more frequent than the a priori

probabilities implied. So the only result of his labors was that he learned that the dice were irregular, and that the a priori probabilities he had assumed were incorrect.

As a mathematical exercise, the dice experiment was a failure, but as an investigation into the properties of his dice, it was a reasonable way to behave. It was also analogous to the way many physical experiments have proceeded. The physical systems we study are composed of many basically identical objects, such as molecules, as alike as casts of the same die. Each molecule can be in one of a variety of states, as the die can show one of several faces uppermost. A collection of many molecules is like a set of many casts, and the sets of results can be related to the properties of an individual molecule or to an individual cast by the mathematics of statistical theory. Distribution functions of physical properties give us clues to the structure of matter, just as distribution functions of gambling results show us some aspects of the game.

There are still stronger parallels between physical systems and games of chance, and those parallels are suggested by the similarities in the distribution functions. We can begin to develop an insight for the parallels by exploring some qualitative features that games of chance have in common. Coin-tossing and dice are not the only games of chance, and other games have their distribution functions as well. Allowing for changes in scale, and sometimes a regular distortion, they all show some strong resemblances. It is not very surprising that they should resemble each other in a crude way; after all, every game has a normal, or most expected result, and the rest of the results are less and less likely, so that they each have a curve with some sort of peak, with decreasing heights for the results progressively different from the norm. But many resemble each other in more detail than that; these have the same mathematical form, called the binomial distribution.

The binomial distribution is basic to all games of chance. Their fundamental similarity is due to the nature of pure chance, in which every event is alike, and no event is influenced by those that come before or after. Each time the dice are thrown, the chance of a particular result does not depend on whether that number has been appearing unusually often. Since each result is independent, there is no way for a run of luck to prolong itself, and the run will continue or cease according to the

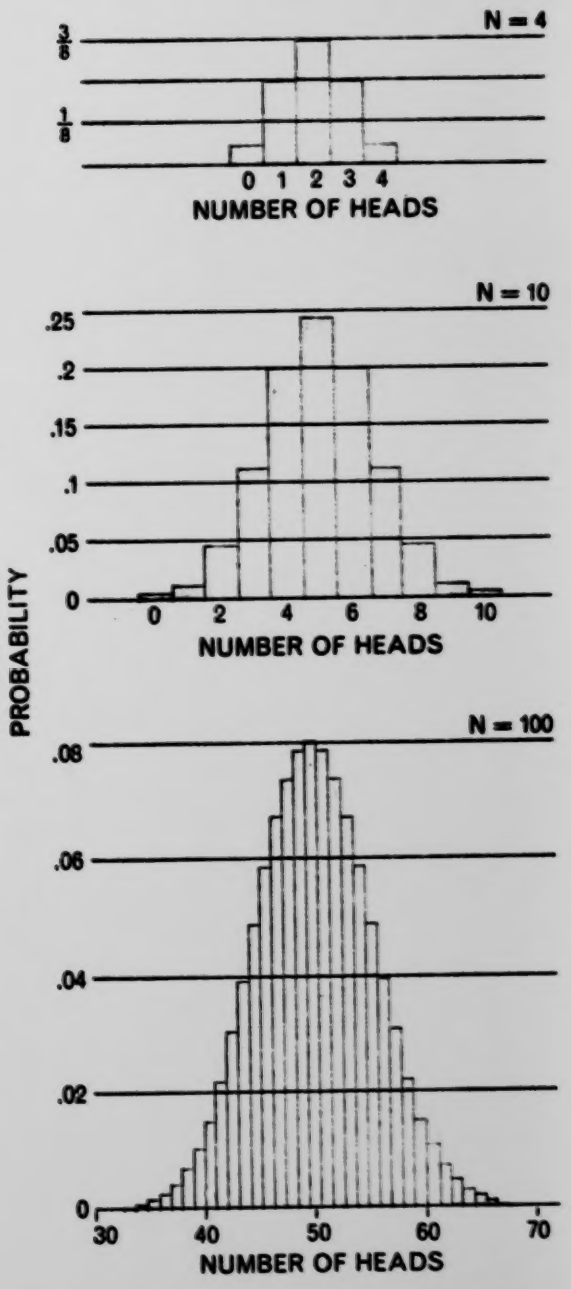


Fig. 2.1 Probability diagrams of the number of heads appearing in a sequence of tosses of a well-balanced coin. N is equal to the number of tosses in each sequence.

results of each successive throw, having the same chance as any other cast. All games of chance are this way, and if we allow for variations in their designs, such as the number of possible results of a single event, they have similar distribution functions. The basic distribution is the binomial; under certain conditions, it takes on forms which are easier to use for computations. The two derived forms most commonly used in physics are the Poisson and the normal (or Gaussian), distributions.

2.2 PRECISION MEASUREMENTS

How can it be that these mathematical forms have a place in physics? Science and gambling seem worlds apart; one representing reason, and the other, the complete lack of it. Is it only an accident that their mathematics show some resemblances? Actually, it is no accident, since there are many ways in which chance plays an important role in science, and when it does, its signature is found in the presence of one of the characteristic distribution functions.

When they do appear, it is often in the most intentional and logical operations. For example, one of the most precise technical manipulations one can do today is to measure a physical quantity. The measurement of the length of a rod, if repeated several times, will give a set of slightly different answers. Of course, more precise instruments will give answers which cluster together more closely: Measurements taken with a micrometer will show less scatter than a series obtained with a ruler. Each set of measurements will have an average, the "length" of the rod, and each set will contain values which become less numerous as they move further from the average. If the measurements are done without systematic errors (such as squeezing the rod too hard with the micrometer or always reading the ruler at a slant), both sets of measurements give the same average value. That is, the ruler values may average to 0.69 inch, say, and the micrometer values average to 0.6883 inch: They agree, if we allow for the lower precision of the ruler. The ruler values have a wider spread, and we may find that several are greater than 0.71 inch, while there are no micrometer values greater than 0.690 inch. The spread, or dispersion, of "wrong" results varies according to the instrument and to the way it is used, but allowing for the difference in precision, both sets of results tend to have very similar shapes. They are usually approximately Gaussian, at least over their central regions, with only an average value and a dispersion to differentiate one set from another. And now, there are no parameters left which might refer to the nature of the rod or the instrument with which the data were obtained.

The Gaussian distribution is good not only for measurements of the length of a short rod by rulers and micrometers, but for any lengths and distances, and for instruments of any precision. There is not even a parameter to describe what sort of physical quantity is being measured; the statistics of "wrong" answers is quite independent of the nature of length. It is just as true of mass, and the preceding description could be immediately extended to describe a measurement of mass simply by changing a few words: "mass" for "length," and perhaps "spring scale" for "ruler" and "analytical balance" for "micrometer." And it is true of all physical measurements, no matter how sophisticated the experiment. The precision of some physical measurements is now greater than 1 part in 10 billion: it is just as true for them. Of course, there are limitations on such a general rule. There must be many possible "wrong" answers from the result of a single measurement for the distribution to be Gaussian. Otherwise, the distribution is composed of discrete possibilities - the binomial distribution is more appropriate in that case, as it is for coin-tossing and

dice. The errors must also be truly random, with no especially large or unusually probable errors of a particular kind or magnitude. "Random" is almost a definition of the pattern of errors that follows the simple statistical distributions. We can interpret the mathematical similarity between the statistics of physical measurements and games of change in two ways: in one sense, a crapshooter is making a series of measurements to establish the average behavior of his dice; in another sense, it is a gamble that random variations will cancel out when we make a physical measurement. Actually, the signature of one of these distributions is usually highly prized by the research worker or statistician, because it indicates that the data has probably been collected without bias: even the "bad" results were kept.

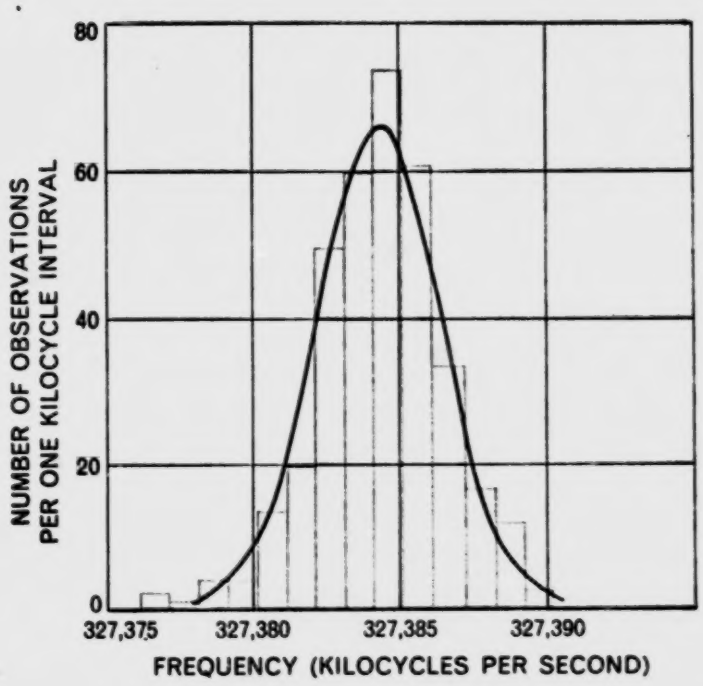


Fig. 2.2 Hyperfine structure separation of the ground state of deuterium, measured by atomic beam resonance by A. G. Prodell and P. Kusch (Phys. Rev. 88, 184 (1952)). Histogram showing the distribution of 350 measurements of the frequency, together with a Gaussian error curve.

2.3 QUALITY CONTROL

But who is to tell us what is "bad"? The variations belong to all aspects of the measurement: the instrument, its use, and the measured quantity itself. They spring from causes ranging from the most trivial, such as careless work, to the most complex, such as are due to the nature of the thing measured. All of these variations contribute to the total spread of the measurements, and there is no way of examining a single set to determine what variations are due to a poor instrument or are characteristic of the thing measured. For the single set of many experiments, all of the random scatter is real, and there are no "wrong" measurements. (Not all variations are random, however - sometimes unusual events interfere with the course of a measurement, and cause some values to deviate from the norm by much larger amounts than expected. When this happens, the deviant result may be tested against a standard distribution, and it may be discarded if it fails to meet quantitative criteria.) Sometimes it is apparent that most of the scatter of results is due to a poor instrument, and that the measured quantity is much more uniform than the data indicated. But unless the measurements are repeated with a more precise instrument, there is no way of reading coarse data to tell which errors are instrumental. However, if more coarse measurements are made, their average becomes more reliable, and in the limit of an infinite number of measurements, it is unaffected by the random errors of the instrument and its operator.

If, instead of measuring the length of a single rod, we measure the lengths of many apparently identical rods, the measurements show wider variations. The variations can be attributed to real variations in the lengths, independent of the experimenter or his micrometer. Variations can be perceived even among supposedly identical rods made by the same automatic machine. Whatever errors contribute to the variations are errors in the way the machine repeats its manufacturing cycles.

These variations, which are due to many imperfections, produce irregular products whose properties are often found to fall on Gaussian distribution curves. No matter what property is considered, or how standardized the product, examination usually discloses similar distributions. They come from the same fundamental behavior of random variations as those of repeated measurements, except that in these cases, one cannot attribute the variations to human mistakes. They are also a feature of automation, and the statistical analysis and limitation of the resulting imperfections form the basis for the fields of "quality control," "process control," or "industrial statistics."

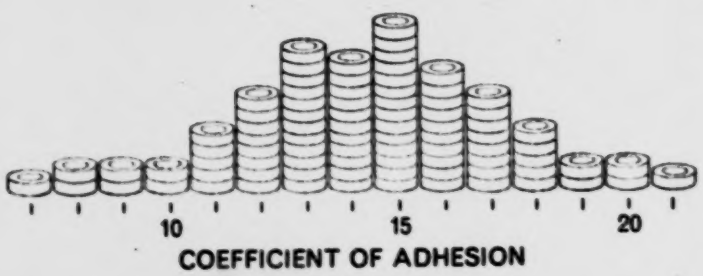


Fig. 2.3 85 rolls of tape classified by adhesion values. After John D. Heide, Industrial Process Control by Statistical Methods (McGraw-Hill Book Company, New York, 1952).

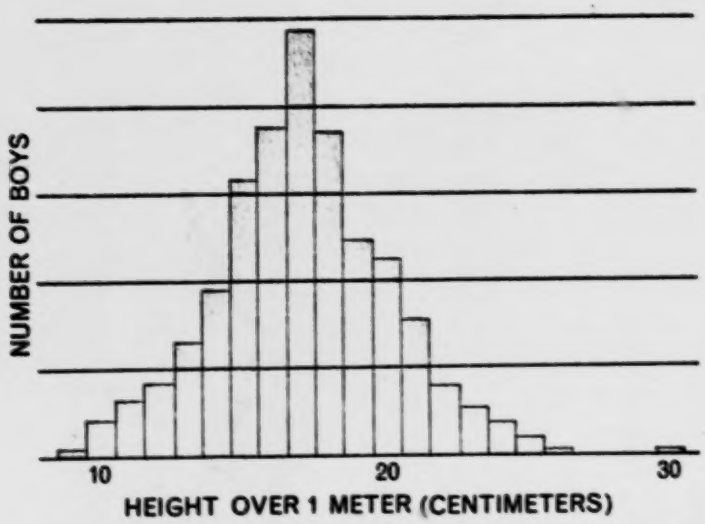


Fig. 2.4 Heights (excess over 1 m) of 767 six-year-old boys, in a group selected for uniform arm length. After Paul Peach, An Introduction to Industrial Statistics and Quality Control (Edwards and Broughton Co., Raleigh, 1945).

In many cases the manufacturing process can be seen to be imperfect. A shaft may have worn bearings, a cutting tool may chatter against the work, or a switch contact may be dirty. When such defects are corrected the machine makes products of greater uniformity. But variations will still be present, due to minor defects in the machine, and these variations will still fit a distribution curve, although one of smaller dispersion. The process of machine improvement is a never-ending spiral toward smaller variations and a narrower range of results, but they are still describable by a distribution.

Somewhere along the line of decreasing errors one finds a limitation in the nature of the materials of the machine and the product. For example, one of the technical properties of metals and alloys is their ability to withstand repeated bending and flexing. The bending produces a network of crystalline dislocations, microscopic cracks, and eventually fracture. Different alloys have enormously different resistances to this sort of fatigue, and their resistances are obviously a matter of great technical importance. In many schools and laboratories there are machines which test standard samples by bending them repeatedly until they break. They work under controlled conditions, with fixed angle of bend, radius of curvature, number of bends per minute, and so forth. With every significant variable kept uniform, apparently identical samples have markedly differing lifetimes: The lifetimes fall on a distribution curve of considerable width. There seems no way to make the pieces break after an identical number of cycles: the scatter is in the nature of the samples, in the random processes of the atomic motions and arrangements that lead to the microscopic dislocations and cracks. If there were a way to make them all show fatigue and crack at the same point, one would be able to save considerable sums of money by replacing parts just before they were able to crack. Just such a supposition, that a

good metallurgist can predict when particular pieces (aircraft wings), were due to break, formed the central theme of a popular novel and movie. It was good entertainment, but terrible physics. And yet, it is almost reasonable to expect that all of the scatter can some day be eliminated. Years ago, variations among samples were much greater. Looking back, we know that they were caused by gross variations in alloy composition, heat treatments, and machining. Modern metallurgical practice has placed all of these variables under much stricter control, and the distribution functions of the alloy properties have responded by becoming much narrower. It is now difficult to make them still narrower, since the relatively easy controls have already been applied. If very great improvements are still to be made, it seems unlikely that they will be accomplished simply by improving our present manufacturing techniques, but will have to depend on quite novel and sophisticated procedures, which will control even the pattern and density of crystalline imperfections in each sample. Until then, it seems that the variations among samples can still be classed as accidental errors in their fabrication, either directly due to the people who design and control the methods and machines, or to imperfections in the machines themselves. Gaussian and similar distributions of the properties of the samples could then arise in essentially the same way they do in a series of physical measurements, out of the randomness of errors and cancellations of their effects.

2.4 BIOLOGICAL VARIATION

If the distributions in the properties of artifacts were only due to human or mechanical errors, we should expect to see very different patterns among natural objects. There are many types of essentially identical things that have natural, or nonhuman origins. Some of the most familiar are living forms of the same species. They are as alike as peas in a pod. How about peas? Mendel used them for his studies of inheritance. The regularities he discovered are expressions of the multinomial distribution, an extension of the binomial for the case in which more than two results of a trial are possible. The statistics of pea colors pointed toward the mechanism of inheritance, and we now understand how the random encounters of pairs of chromosomes lead to the regularities of the Mendelian laws. Within a single genotype there are remaining variations in other characteristics: size, weight, firmness, or germination time, and these variations are usually Gaussian. As it is for peas, the same distribution law occurs throughout the biological world; no form of life is exempt, not even our own. There are tons of statistics on human measurements, ranging from the ordinary sort of height, weight, and so forth (see Fig. 2.4), to microscopic, physical, and chemical studies requiring highly specialized instruments. Whenever a large number of samples is obtained, the familiar mathematical forms emerge. Normal distributions characterize nonphysical qualities as well, such as I.Q. and test scores. When a class is graded on a curve, that curve is usually an approximation to one of the simple distribution laws. In fact, the branch of mathematics called statistics has origins closely associated with the analysis of human populations. Among the early studies were the proportion of a population that could be expected to bear arms, life insurance tables, and the immunity to smallpox. Sometimes it is easy to understand how a particular phenomenon, which has its basis in accident, leads to the statistics of chance. For example, an analysis of the number of deaths in given years by kicks from a horse disclosed a Poisson distribution. Yet, the same distribution describes the rate of telephone calls arriving at a central switchboard. And even the apparently intentional and organized actions of society

seem to follow the same pattern: the numbers of outbreaks of war per year from 1820 to 1939 have been found to follow a Poisson distribution remarkably well.

Observations of binomial, Poisson, and Gaussian and related distributions in biological populations is one thing, but understanding why is another. It is not at all obvious that the "Law of Error" should also describe the variations among biological properties. Its interpretation caused a controversy among scientists and mathematicians that lasted over a century. The arguments are worth describing here, because some of them help to illuminate the mystery of the same law's appearance in physics.

In 1835 the Belgian astronomer Quetelet published "Essai de Physique Sociale," with which he gave a substantial basis to the new theory of statistics. He was the first to present convincing arguments for what has been called the constancy of large numbers, with particular application to men. The success of Quetelet's statistics combined with the contemporary rationalism to convince him that he had demonstrated that society obeyed eternal and natural laws. As the measurements of a physical quantity were scattered about the true value because of error, so the ideal biological models were thought to be imperfectly copied by nature, which, in his view, always strives for perfection. Quetelet's ideas were comfortable for his society and his time: It was an age of supreme reason and natural perfection. But as the philosophy of society relaxed to one of greater liberalism, new interpretations were made. In the debate, which still echoes, the distinctions between "error" and "variation" have been blurred.

In physics we are on ground further removed from ideas of perfection and purpose, errors, and mistakes. The familiar distribution laws seem to be everywhere. The velocities of the atoms of a gas are found to be Gaussian in any direction; so are the velocities of the conduction electrons in a semiconductor. The numbers of atoms of a radioactive bit of matter that decay in each second follow a Poisson distribution law; so does the number of molecules of hydrogen in each cubic centimeter of interplanetary space. Instead of lengthening this list of examples from physics, we will end the discussion of the statistical laws here, to examine their basis by mathematical analysis.

3 AN INTRODUCTION TO THE MATHEMATICS OF PROBABILITY

Coin-tossing is one of the simplest examples of systems of random events, and yet it can lead us to an understanding of important and current fields of physics. The theory is readily adaptable to phenomena as diverse as the magnetism of solids, the evaporation of liquids, and the structure of alloys. These phenomena share a reliance on microscopic units which can be idealized as having only two possible states of conditions of some sort: ↑↓, + −, ○ ●, or heads-tails. By simple extensions of the mathematics, it will be possible to analyze phenomena springing from more than two states, and these will include the molecules in a gas that we set out to discuss in Chapter 1.

3.1 THE BINOMIAL DISTRIBUTION: A SYMMETRIC COIN

We assume that the probability of coming up heads is equal to the probability of tails:

$$p(H) = p(T).$$

A result is certain to be either heads or tails, so that for a perfectly balanced coin,

$$p(H) + p(T) = 1;\ p(H) = p(T) = 1/2. \tag{3.1}$$

For a sequence of several tosses, our primary rule is that each toss is an independent event, unaffected by preceding results, and always with the same probabilities, no matter how long we play.

Yet, in a sequence of throws, we know that it is most likely that heads and tails will appear about equally often, and that it is quite unlikely that only heads will come up. The reason for this is not that the coin or luck eventually makes a correction, but is simply a result of the independence of probabilities of each toss. For instance, let us suppose that there has been a sequence of tosses, perhaps 100 heads in a row, and that the probability of this occurrence is some (small!) number p(100H). Now, when the next toss is made, the probability of one more head is 1/2, just as for a single toss. This means that, if we have been so lucky (or unlucky) as to get 100 heads in a row, we have only one chance in two of extending our string one more toss, to 101 in a row. We can then write that the probability for 101 heads is

$$p(101H) = p(H)\ p(100H) = \tfrac{1}{2}p(100H).$$

Now, if the toss <u>is</u> a head, we know what the chance is of extending the string to 102:

$$p(102H) = p(H)\ p(101H) = \tfrac{1}{4}p(100H)$$

Thus, we see that the chance of extending the string of successes decreases with each additional success, by the same factor each time. The improbability of a long string is just the result of the repeated product. Therefore, the probability of 100 heads in a row is the repeated product of 100 factors of $\frac{1}{2}$ or

$$p(100H) = (\tfrac{1}{2})^{100}.$$

Of course, the probability of 100 tails in a row is the same. Moreover, since the probability of a tail on any single toss is precisely the same as the probability of a head, we get the same chance for a string of 99 heads in a row, and then one tail, as for 100 heads. Or of 23 heads, a tail, and 76 heads. Or of 23,heads, and then 77 tails. Or even of an alternating set of

heads and tails, adding up to 50 heads and 50 tails! This seems wrong, until we realize that we are specifying precise results for individual tosses, instead of simply letting the vagaries of chance work things out in the long run. When we relax our conditions, and let heads or tails come up on any individual toss, we find that the string is likely to have about 50 heads and 50 tails. But we have to include runs in which we may get clusters of several heads or tails in a row, as well as alternating sides, and also strings in which there are first 50 heads and the rest tails. We begin to see that the dominance of the average behavior is that there are many ways of reaching it, as long as we do not specify the detailed route.

For N tosses, each distinct sequence has probability $(1/2)^N$. The generation of each sequence can be seen as a branching diagram, as the Fig. 3.1 below: At each junction, the chance of following one of the two branches is 1/2, so that the chance of taking one particular route containing N junctions is $(1/2)^N$. How many different routes are there for a diagram of N junctions? Since each junction has two branches, there are simply 2^N different routes. Since the probability of each route is $(1/2)^N$, the sum of probabilities of all routes is therefore $2^N(1/2)^N = 1$, as it should be: If N junctions have been passed, it is certain that we have taken one of the possible routes to the end.

Usually, we are not interested in the order in which heads or tails appeared, but simply in how many there were in the whole sequence. If we put the different sequences into such classes, a distribution begins to emerge, with some classes much more popular than others: examples of such distributions are shown in Fig. 3.2. If the number of tosses is very small, we can sort the sequences into classes by preparing tables, but tables are impractical for longer runs - the number of entries for N = 20 would be $20(2)^{20}$ = 21,171,520. Clearly, some shortcuts are necessary. Fortunately, they are not hard to develop. Suppose we wish to determine how many different sequences of N tosses contain n heads and m tails in any order. To begin, we think of each H as labeled with a distinguishing mark. Let them be numbered H_1, H_2, ... H_n; they may each appear on any one of the N tosses not already "filled" by another H. After all the H results are distributed in a sequence, the remaining (N − n) tosses are occupied by the m T's that remain: m and (N − n) are the same.

Beginning with the first, we can imagine H_1 in any of N places. Since one space is filled, there are only (N − 1) openings for H_2, and progressively fewer for each additional one, down to (N−n+1) for H_n. All together,

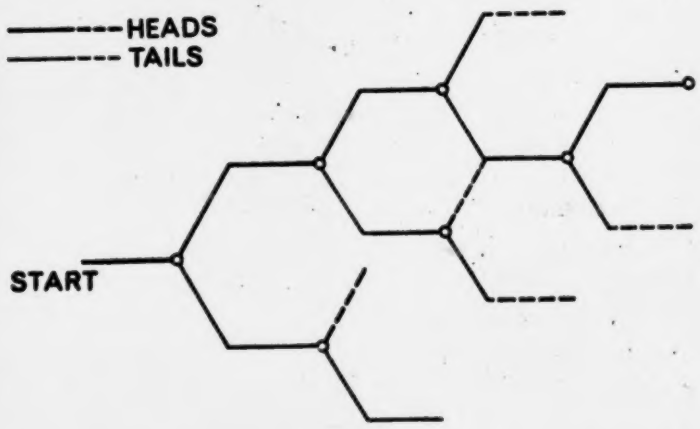


Fig. 3.1 Alternative sequences of the results of tosses of a coin.

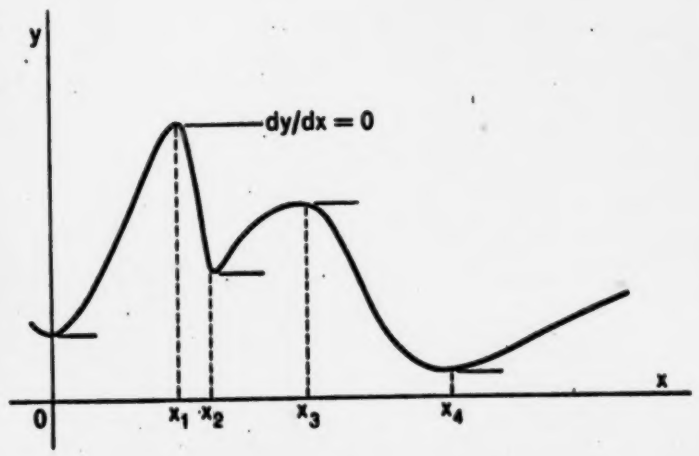


Fig. 3.2 Finding maxima and minima of functions by differentiation.

the total number of combinations is the continued product

$$N(N-1)(N-2)\ \ldots\ (N-n+1).$$

A convenient symbol for a continued product is the factorial, defined for an integer z by

$$z! = z(z-1)(z-2)\ \ldots\ (3)(2)(1),$$

with 0! = 1. The number of different arrangements of n distinguishable H in N tosses can then be written as $N!/(N-n)!$. But in an actual sequence the different H are not distinguishable, and there is no difference between a run such as

$$H_1\ T\ T\ H_2\ T\ H_3\ \ldots$$

and another of the form

$$H_3\ T\ T\ H_1\ T\ H_2\ \ldots$$

The factor by which we have overcounted is the number of different orderings we can make of n distinguishable things: it is simply n!. Therefore, the number of different sequences of n identical H in N tosses is $N!/(N-n)!n!$. This number is the statistical weight w(nH, mT) of the category of n heads and m tails in N = n + m tosses: It is an important result, and we rewrite it in more symmetric form:

$$w(nH,mT) = \frac{N!}{n!m!} \qquad (3.2)$$

Since the probability of every different sequence is $[p(H)]^N$, the probability of getting any member of the class (nH, mT) is

$$p(nH,\ mT) = w(nH,\ mT)[p(H)]^N. \qquad (3.3)$$

The idea of a statistical weight is useful in physics, when it is practically impossible to distinguish individual arrangements of microscopic quantities, and only those properties characteristic of classes of arrangements might be determined.

The graphs in Fig. 2.1 were drawn according to Eqs. (3.2) and (3.3). Additional properties of the symmetric binomial distribution are described in the next section.

Among the most interesting features to explore are the expected number of heads and the dispersion. We expect that the distribution should have a shape that is a progressive development from the discrete graphs of Fig. 2.1, with steps becoming less distinct and the whole curve narrowing as N increases. Instead of graphing a succession of curves, we can discover these features by analysis.

For numbers larger than 10 the factorial is tedious to calculate; (however, there are published tables of z! for $0 > z > 1000$). In statistical physics it is almost always sufficiently accurate to use an approximate expression for z! when $z > 10$. Stirling's approximation is

$$z! = (2\pi z)^{\frac{1}{2}}\left(\frac{z}{e}\right)^z, \qquad (3.4)$$

where e is the base of the natural logarithms. The formula is accurate to within (9/z) percent. Using Stirling's approximation for the factorials in Eq. (3.2), we obtain a more convenient form for the statistical weight of the class of n heads and m tails:

$$w(nH,\ mT) = \left(\frac{N}{2\pi nm}\right)^{\frac{1}{2}} \frac{N^N}{n^n m^m} \qquad (3.5)$$

It is convenient to write the expression for statistical weight in terms of a single parameter for the fraction of heads in the total number of trials. Defining the ratio r = n/N, then (1 − r) = m/N, and substituting in Eq. (3.5), the statistical weight can be expressed as a function of r and N only:

$$w(r,N) = [2\pi Nr(1-r)]^{-\frac{1}{2}} \frac{1}{r^{rn}(1-r)^{(1-r)N}}. \qquad (3.6)$$

To locate the value of r corre-

sponding to the greatest statistical weight, we search for the region where a small change in r does not affect the result. It is the same principle that one could use to locate the peak of a physical mountain, by finding a spot where a little motion forward or backward would not cause any change in one's altitude. Of course, the technique would give all the intermediate peaks and valley floors as well, so it must be used in conjunction with a few rough observations in order to select the highest value from among all the flat spots. But it should give us little trouble in the present application, since we expect to find only one peak, with steadily falling values on either side.

When N is sufficiently large, the variable r can be treated as continuous rather than discrete. Instead of trying to locate the peak by the method of finite differences, we can then use the differential calculus, which is much easier (as long as one knows differential calculus!). For any continuous and differentiable function $y = f(x)$, the extrema (maxima and minima) of y correspond to places where the derivatives of y with respect to x are zero: There is an extremum at $x = x'$ if

$$\left(\frac{dy}{dx}\right)_{x=x'} = 0.$$

In the Fig. 3.1, there are maxima at $x = x_1$ and x_3, and minima at $x = 0$, x_2, and x_4. By differentiating f(x) and setting it to zero, we should find all of these values as roots of the equation. But we must inspect y itself to determine that the root $x = x_1$ is the one that yields the largest value of y.

Applying the technique to Eq. (3.6) we can find r_m, the most probable value of r, which corresponds to the most frequently appearing fraction of heads. We could follow the prescription by differentiating p(r,N), directly, as

$$\frac{dp^{(r,N)}}{dr} = \frac{dw^{(r,N)}}{dr}\,[p(H)]^N = 0,$$

for the case p(H) = p(T).

Instead of working with the probability directly, it is much easier in the case of large exponents to work with the logarithm of the probability. In this way we will be able to separate the individual terms and to pick out those that are most important. The habit of working with logarithms of functions rather than the functions themselves is characteristic of statistical physics; it will come up again. Since ln p is a monotonic function of p, the extrema of ln p occur at the same positions as the extrema of p itself.

Remembering that $\ln(ab) = \ln a + \ln b$, $\ln(a/b) = \ln a - \ln b$, and $\ln(a)^b = b \ln a$, taking the logarithm of both sides of Eq. (3.6) gives us

$$\ln w(r,N) = -\tfrac{1}{2}\ln r - \tfrac{1}{2}\ln(1-r) - r\,N \ln r - (1-r)\,N \ln(1-r) + N \ln p(H).$$

Differentiating with respect to r, and using $d/dx\,(\ln x) = 1/x$, we obtain

$$\frac{d}{dr}\left[\ln p(r,N)\right] = \frac{1}{2}\left[\frac{1}{(1-r)} - \frac{1}{r}\right] + N \ln\left[\frac{(1-r)}{r}\right].$$

If N is very large the second term is dominant in the intermediate range of r. In this region the only point where

$$N \ln\left[\frac{(1-r)}{r}\right] = 0$$

is that for which $(1-r)/r = 1$. Thus, $r_m = \frac{1}{2}$. The result is independent of N: The most probable result is that heads comes up half the time in any length of sequence.

Sometimes the average value in a distribution is not the same as the most probable value. They will be different when the values are not distributed symmetrically about the peak of the curve (the distribution is skewed).

For example, if in a group of 100 test papers there are 10 perfect scores but the rest are distributed in steps of 1 from 1 to 90, the most probable score is 100, but the average is only 50.95. The average value of x, which is usually written $\langle x \rangle$ or $\bar{x}$, can be found from the probability distribution by remembering that the probability is equal to the fraction of times that a result appears. The average is formed by totaling all of the values that appear and then dividing by the number of entries. Thus, if there are N trials, a probability $p(x_1)$ means that N $p(x_1)$ results gave the value x_1. These add the contribution x_1 N $p(x_1)$ to the total. The average is then

$$\langle x \rangle = \frac{1}{N}[x_1 N p(x_1) + x_2 N p(x_2) + \ldots$$

$$x_N N p(x_\nu)]$$

$$= \frac{1}{N}\sum_{i=1}^{N} x_i N p(x_i) = \sum_{i=1}^{N} x_i p(x_i). \qquad (3.7)$$

If the probability varies little from one value of x to the next, so that $p(x_{i+1}) - p(x_i) \ll p(x_i)$, we may treat x as a continuous variable. In this case, the probability dp(x) that the result lies between x and x + dx is proportional to the interval dx:

$$dp(x) = f(x)dx, \qquad (3.8)$$

f(x) is known as the probability density. It is a normalized function, since the probability is unity that x has some value out of the total possible range:

$$\int dp(x) = 1 = \int_{-\infty}^{\infty} f(x)dx \qquad (3.9)$$

Averages of functions of variables are often of interest. For example, if results having large values are particularly important, we may want to develop some property of the distribution which emphasizes the large results more than the smaller ones. If a high school curriculum is designed to stimulate the few unusually talented students, their grades may be improved to very high levels, but they would make little difference to the class average. We could emphasize the appearance of their high scores by averaging not the scores themselves but the squares of the scores, or even higher powers. The prescription for averaging the squares of quantities is formally the same as for the quantities themselves.

$$\langle x^2 \rangle = \sum_{i=1}^{\nu} x_i^2 p(x_i) \qquad (3.10)$$

for discrete values of x, and

$$\langle x^2 \rangle = \int_{-\infty}^{\infty} x^2 f(x)dx$$

for continuous values.

For any integrable function g(x), the prescription is

$$\langle g(x) \rangle = \int_{-\infty}^{\infty} g(x)f(x)dx. \qquad (3.11)$$

A very interesting feature of the averaging process is that the average of a function is not identical to the function of the average: they are noncommuting operations in general. A simple example is x^2, and the set of numbers in Table 3-1 illustrates the inequality $\langle x^2 \rangle \neq \langle x \rangle^2$.

	x	x^2
	−3	9
	−2	4
	−1	1
	0	0
	1	1
	2	4
	3	9
TOTALS:	0	28
AVERAGES:	$\langle x \rangle = 0$	$\langle x^2 \rangle = 4$

Table 3.1 An illustration of the noncommutability of the operations of averaging and of making an arbitrary function of a variable. The example shows that $\langle x \rangle^2 \neq \langle x^2 \rangle$.

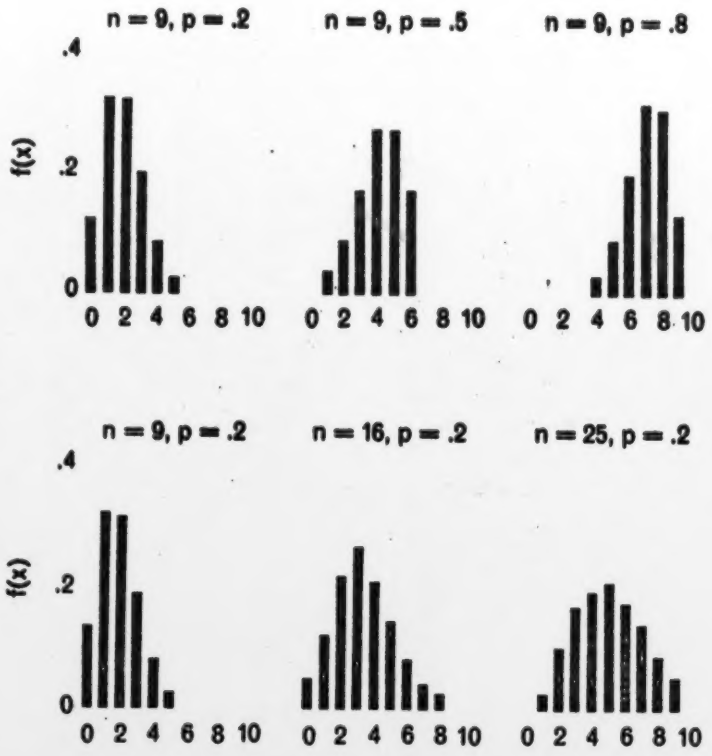


Fig. 3.3 Binomial distributions for various values of N and p. After G. P. Wadsworth and J. G. Bryan, Introduction to Probability and Random Variables (McGraw-Hill Book Company, New York, 1960).

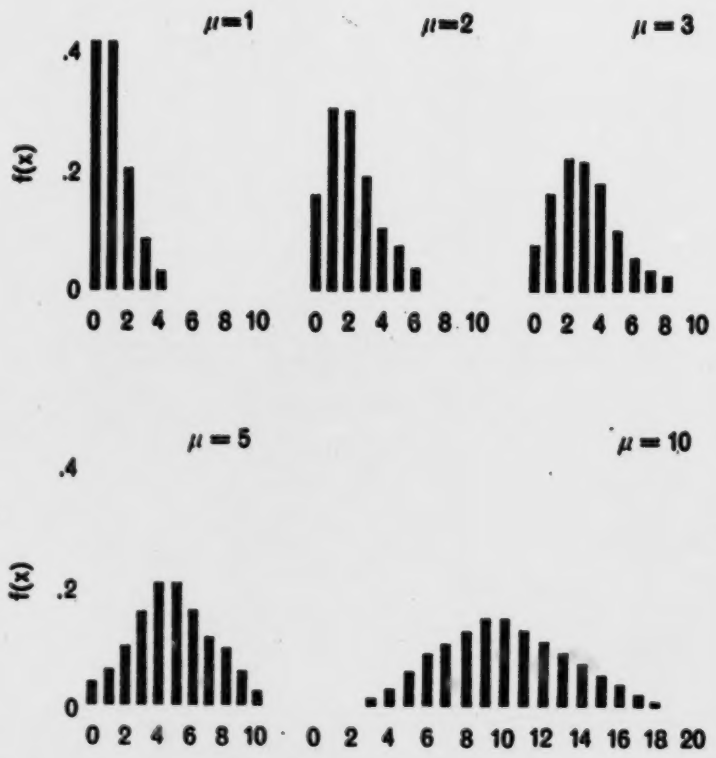


Fig. 3.4 Poisson distribution for various values of μ. After G. P. Wadsworth and J. G. Bryan (op. cit.).

The relative widths of distributions can be expressed in terms of $\langle x^2 \rangle$ and $\langle x \rangle^2$. The dispersion σ is defined by the relation

$$\sigma^2 = \langle x^2 \rangle - \langle x \rangle^2. \qquad (3.12)$$

It can also be written in the form

$$\sigma^2 = \langle (x - \langle x \rangle)^2 \rangle. \qquad (3.13)$$

For the binomial distribution, the dispersion can be calculated from the discrete finite series. For a symmetric coin, the result is that the dispersion in the number of heads appearing in sequences of N tosses is

$$\sigma_n = \tfrac{1}{2}\sqrt{N},$$

showing that the magnitude of the deviation from an equal number of heads and tails increases as the sequence gets longer. But the deviations do not increase as rapidly as the sequences themselves. The relative deviation of n decreases: if we express the result in terms of the ratio $r = n/N$, the dispersion in r is

$$\sigma_r = \langle r^2 \rangle - \langle r \rangle^2 = \frac{1}{N^2}\,\sigma_n^2.$$

Hence

$$\sigma_r = \frac{1}{2\sqrt{N}},$$

showing that the relative distribution gets narrower as N increases. In the limit of an infinite number of trials, the curve has no width at all, giving the precise result $r = \tfrac{1}{2}$. In the limit, we return to the way we defined the probability for each toss, by assuming an infinite number of trials.

3.2 OTHER DISTRIBUTIONS

3.2.1 The Skew Binomial Distribution: A "Plugged Nickel"

If the probabilities of the two possible results A and B of a trial are different, the binomial distribution is no longer symmetric. Suppose

that the probabilities are p for the A result and q for the B, with $p + q = 1$. By a simple extension of the development of the symmetric binomial distribution, we find that the probability for n results of type A in a sequence of N trials is

$$p(n) = \frac{N!}{n!(N-n)!} p^n q^{N-n}. \qquad (3.14)$$

The distribution is skewed for all cases in which $p \neq q$, as seen in several cases graphed in Fig. 3.3.

3.2.2 The Multinomial Distribution

If there are more than two mutually exclusive results possible from a single trial, the probabilities of combinations are describable by a multinomial distribution. Let the probability for a type 1 result be p_1, for a type 2 result be p_2, and so forth, the total number of different results being ν. By an extension of the development of the binomial distribution, the probability for n_1 results of type 1, n_2 of type 2, etc., to occur in a sequence of N trials is

$$p(n_1, n_2, \ldots n_\nu) = \frac{N!}{n_1! n_2! \ldots n_\nu!} p_1^{n_1} p_2^{n_2}, \ldots p_\nu^{n_\nu}, \qquad (3.15)$$

where

$$p_1 + p_2, +, \ldots p_\nu = 1.$$

3.2.3 The Poisson Distribution

This is a limiting form of the binomial under certain conditions. The conditions are that $p \to 0$ and $N \to \infty$ in such a manner that the product Np tends to a constant limit μ. The Poisson distribution for n successes is

$$p(n) = e^{-\mu} \frac{\mu^n}{n!}. \qquad (3.16)$$

Several examples of Poisson distributions are shown in Fig. 3.4; notice the similarities between Fig. 3.4 and the skewed binomial distributions in Fig. 3.3. The correspondence between the Poisson and binomial distribution in the region of small p and large N is also evident in Table 3.2, which compares the two for the case of $p = 1/500$ and $N = 500$.

n	BINOMIAL	POISSON
0	.3675	.3679
1	.3682	.3679
2	.1841	.1839
3	.0612	.0613
4	.0153	.0153
5	.00303	.00307
6	.00050	.00051
7	.000071	.000073

Table 3.2 Comparison of binomial probabilities for $p = 1/500$ and $N = 500$, with Poisson probabilities for $\mu = Np = 1$.

3.2.4 The Gaussian Distribution

The Gaussian distribution is also a limiting form of the binomial: It is appropriate for the case of large N, such that the values of results can be treated as a continuous variable x. In terms of the product $\mu = Np$ and the dispersion $\sigma = \sqrt{Npq}$, the probability

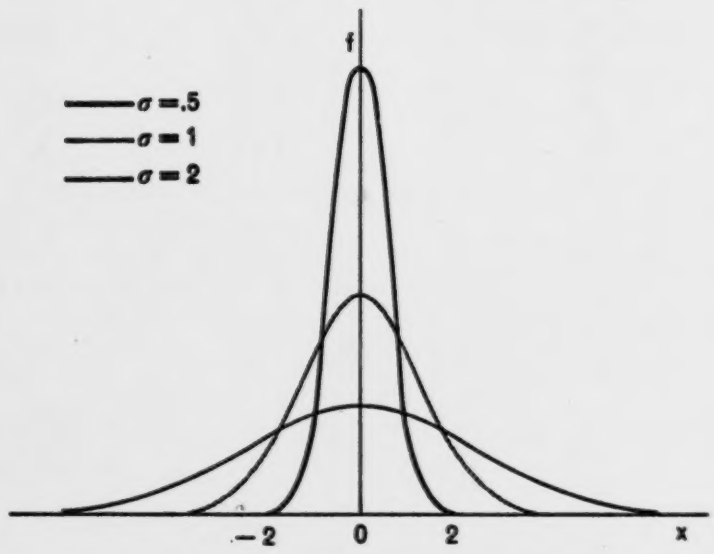


Fig. 3.5 Gaussian distributions for $\mu = 0$ and various values of σ. After G. P. Wadsworth and J. G. Bryan (op. cit.).

that the number of results lies between x and x + dx is

$$dp(x) = f(x)\,dx$$

$$= \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] dx. \tag{3.17}$$

The Gaussian distribution is usually a satisfactory approximation for the binomial for $Np > 5$. It also approximates the Poisson distribution when μ is sufficiently large. The Gaussian distribution is particularly useful because it has convenient mathematical properties; it is generally used in place of the Poisson and binomial distributions whe whenever it is reasonably accurate to do so. In Fig. 3.5 several samples are shown for fixed μ and various σ. The dispersion controls the width, while μ controls the position of the curve: Varying μ would cause the sample curves to slide to higher or lower x, but with no change in shape. The formula for the dispersion in the Gaussian distribution is the same as that used in the discussion of the binomial distribution, that is, the value of $\langle x^2\rangle - \langle x\rangle^2$ for the Gaussian distribution is σ^2.

The probability density for any distribution is normalized, according to Eq. (3.9). For the Gaussian distribution we therefore have

$$\int^{\infty} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] dx = \sigma\sqrt{2\pi}. \tag{3.18}$$

Averages of x and x^2 for the Gaussian distributions are given by the definite integrals:

$$\langle x\rangle = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} x \exp\left[\frac{-(x-\mu)^2}{2\sigma^2}\right] dx = \mu,$$

$$\langle x^2\rangle = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 \exp\left[\frac{-(x-\mu)^2}{2\sigma^2}\right] dx$$

$$= \sigma^2 + \mu^2. \tag{3.19}$$

These averages will be useful in Chapter 4.

4 POSITIONS AND VELOCITIES OF MOLECULES OF A GAS

The statistics of pure chance can be translated into the statistics of several physical systems. One of the simplest examples concerns the distribution of molecules in a gas. If we begin with a question about the distribution in space, the statistics of a symmetric coin can be applied immediately to its solution. Increasingly sophisticated questions about the same molecules can be answered by using distribution functions derived from the binomial law. The exploration will lead us into the elementary kinetic theory of gases.

4.1 THE DISTRIBUTION IN SPACE

Suppose that we have a quantity of gas sealed in a container. How uniformly do the atoms arrange themselves throughout the volume? To make the problem as simple as possible at first, we will consider the container divided into two equal volumes.A and B. We imagine the atoms to be in random motion throughout the total volume, each particle moving back and forth in an irregular path of straight lines by a succession of reflections with the container walls. The roughness of the walls causes the path to vary endlessly, so that an atom threads its way through the container in every direction, eventually coming arbitrarily close to any point, from all angles. No portion of the volume is avoided or preferred (except perhaps for the regions close to the walls), and therefore, the chance of finding the atom within a specified portion of the container is proportional to that fraction of the total volume, independent of the location or shape of the portion. Therefore, if the volume of the container is divided into two equal portions of any shape, the chance that a specified atom is in either portion is 1/2. This makes the probabilities of the atomic locations in the two regions equivalent to the results of the tosses of a symmetric coin. To pursue the analogy, we must guarantee the independence of probabilities of location of the individual atoms, corresponding to the independence of results of individual tosses. To effect this correspondence, we assume that the density of atoms is sufficiently low, so that collisions between them are infrequent: Each atom is rarely disturbed by an encounter with another. Although in this development we imagine atoms as if they only interacted by bumping, real atoms or molecules sometimes have long-range interactions. When two real molecules are within their range of interaction, their positions and velocities are not completely independent: They are found to be correlated. In the present example we consider the average separation to be much greater than the range of interaction, so that their positions and velocities are uncorrelated. Therefore, if two atoms are in the whole container, the probability that both are simultaneously in the A half is 1/4. The chance of finding one (either one), in A and the other in B is 1/2. When N atoms are enclosed, the chance that n are in A is evidently the same as the chance that n out of N tosses of a symmetric coin will be heads, and it is given by the binomial distribution.

In Chapter 3 some properties of the binomial distributions are analyzed, and the results are useful here. They show that the average atomic arrangement will be symmetric for any nu number of molecules: 1/2 in A and 1/2 in B. At any moment, the arrangement may be different from the average, the chance of a certain deviation being equal to that of the same numerical deviation being equal to that of the same numerical deviation in a sequence of

tosses. But the gas is not frozen in its distribution, since the individual atoms continue to travel through the container, and the distribution changes continually. The time for an appreciable change to occur is difficult to calculate with any precision, but we can readily estimate its order of magnitude.

If the container walls are microscopically rough, as almost all real walls are, atomic reflections are diffuse, or random in direction. By such reflections, the spatial distribution of all the atoms is made nearly independent of the previous arrangement as soon as each atom has time to make one collision with a wall. Therefore, the "lifetime" of a particular arrangement in a volume is $L/\langle v \rangle$, where L is a characteristic length of the container, and $\langle v \rangle$ is the average speed of an atom. If instead of concerning ourselves with the arrangement relative to the two halves of V, we ask about the distribution in a smaller part of the volume, the lifetime will be smaller, corresponding to the dimensions of that small part. The lifetimes for the arrangements of common molecules at ordinary temperatures in laboratory-sized volumes are very short. Borrowing a result from a later section of the chapter, $\langle v \rangle$ for low-mass molecules is on the order of 1000 m/sec at room temperature. Therefore, the time for a rearrangement to occur in a 1-m cube is about a millisecond. For this volume, the system passes through a succession of many arrangements rapidly. However, the corresponding lifetime of a spatial distribution in a galaxy of 100,000 light-years in diameter is about 3×10^{10} years. Hence, for small volumes, a series of observations lasting a second may be sufficient in averaging the system over many arrangements, but for most astronomical volumes we can only examine a single arrangement even if we observe over the course of many years.

If either the astronomical or the laboratory-sized samples of gas are in an abnormal arrangement at a given moment, perhaps because of some external influence which has been suddenly removed, they will each lose all memory of that situation in a few characteristic rearrangement times, after which the positions will be in statistical equilibrium, moving from one arrangement to another according to their relative statistical weights. The time for this system of independent particles to relax from an unusual situation caused by a change in its external environment is therefore about the same as the time for new arrangements to occur under steady conditions. The equilibrium time is not always identical to the time for a major reorganization for all types of systems - the relaxation time and the redistribution time are comparable in this example because the walls are assumed to be perfectly rough, the molecules rarely collide, and also because we are now considering only their arrangements in space.

Although every imaginable arrangement is possible, some are much more likely than others, having greater statistical weights. From the results of Chapter 3, we see that the atoms distribute themselves more perfectly between the two halves as their numbers increase. Specifically, the percentage mean deviation from an equal division varies as $1/\sqrt{N}$. For 1 cubic meter of gas at the density of interstellar space, which is about 1 atom/cm^3, the mean deviation is about 0.1 percent. But for 1 cubic meter of air at sea-level pressure, in which there are about 3×10^{22} molecules, the mean deviation is about 10^{-9} percent.

We need not limit ourselves to the case of equal subdivisions of the container. The skewed binomial distribution for an asymmetric coin is appropriate to the arrangements of molecules between two unequal parts of the volume. It is described in Chapter 3, and predicts that the average division of atoms is proportional to the ratio of volumes, according to the a priori probabilities for individual particles. As in the example of two equal portions, the perfection of the ratio of the populations in two unequal portions

is also improved as the total number of particles increases. But if one of the portions becomes so small that the average number of particles it contains is much smaller than the total, then its population will fluctuate by a larger percentage. If, for instance, the sample is so small that the average number it contains is only one atom, it will be subject to fluctuations comparable to the average itself. The larger volume, on the other hand, will not fluctuate at all, since it contains essentially all of the particles, a constant number. Under these conditions, the numbers of particles appearing in the small volume will follow a Poisson distribution. The mathematical form of the Poisson distribution is described in Chapter 3 for the example of a very asymmetric coin, in the limit of a very small probability of heads and very long sequences. For our physical example, the probability of an atom's appearance in the small sample volume is equivalent to the small chance of heads, and the large total number of atoms is equivalent to the large total number of tosses. In terms of the average number $\langle n \rangle$ of particles per unit volume of the whole co container, the probability for finding exactly n particles per unit volume of the small sample is, after Eq. (3.16),

$$p(n) = \frac{\langle n \rangle^n e^{-\langle n \rangle}}{n!}. \qquad (4.1)$$

Some examples of the Poisson distribution are graphed in Fig. 3.4. The first example, with $\mu = 1$, corresponds to the physical example $\langle n \rangle = 1$ discussed above. It shows that there is just as much chance for finding 0 particles as for 1 (e^{-1} for each), half as much for 2, etc. There is thus a very wide scatter in the percentage deviation. For larger $\langle n \rangle$ or μ the absolute width of the distribution increases, but the width relative to the total decreases: In the last example $\mu = 10$, and here the mean deviation is reduced to about 25 percent of the average value. In this case, the distribution is nearly symmetric, and begins to resemble a Gaussian distribution. The approximation becomes quite accurate by the time μ has increased to 50; in Table 4.1 the binomial, Poisson, and Gaussian distributions are compared for this situation. Because its mathematical properties are more convenient for calculations, we prefer to use the Gaussian form when it is appropriate. The Gaussian distribution is a good approximation when n is large enough to be treated as a continuous variable. Translated into terms convenient for the physical case, the Eq. (3.17) for the Gaussian distribution can be translated to the probability f(n)dn that the number of particles per unit volume lies between n and n + dn,

$$f(n)dn = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(n - \langle n \rangle)^2}{2\sigma^2}\right]dn. \qquad (4.2)$$

We have been considering the distribution of molecules as a fluctuating population in a sample small "cell" of the whole container, a distribution in time. We can also consider the distribution from the standpoint of variations in space. In the equilibrium state every small region of the container is as probable a location for an atom as any other cell of the same size. Therefore, there is nothing to distinguish the fluctuations in one

n	BINOMIAL	POISSON	GAUSSIAN
25	.0000	.0001	.0000
30	.0006	.0010	.0007
35	.0052	.0057	.0054
40	.0212	.0205	.0215
45	.0460	.0442	.0458
50	.0569	.0570	.0563
55	.0424	.0442	.0422
60	.0199	.0205	.0201
65	.0061	.0057	.0063
70	.0013	.0010	.0014
75	.0002	.0001	.0002

Table 4.1 Comparison of Gaussian and Poisson approximations to the binomial probability, when N = 2500 and p = 0.02.

cell from those in another, and any cell displays arrangements similar to those expected for any other cell, except that they will occur at different times. Therefore, a description of the instantaneous atomic populations of many cells is equivalent to the description of successive stages of a single cell. Interpreted in this fashion, Eqs. (4.1) and (4.2) are descriptions of the distribution in space.

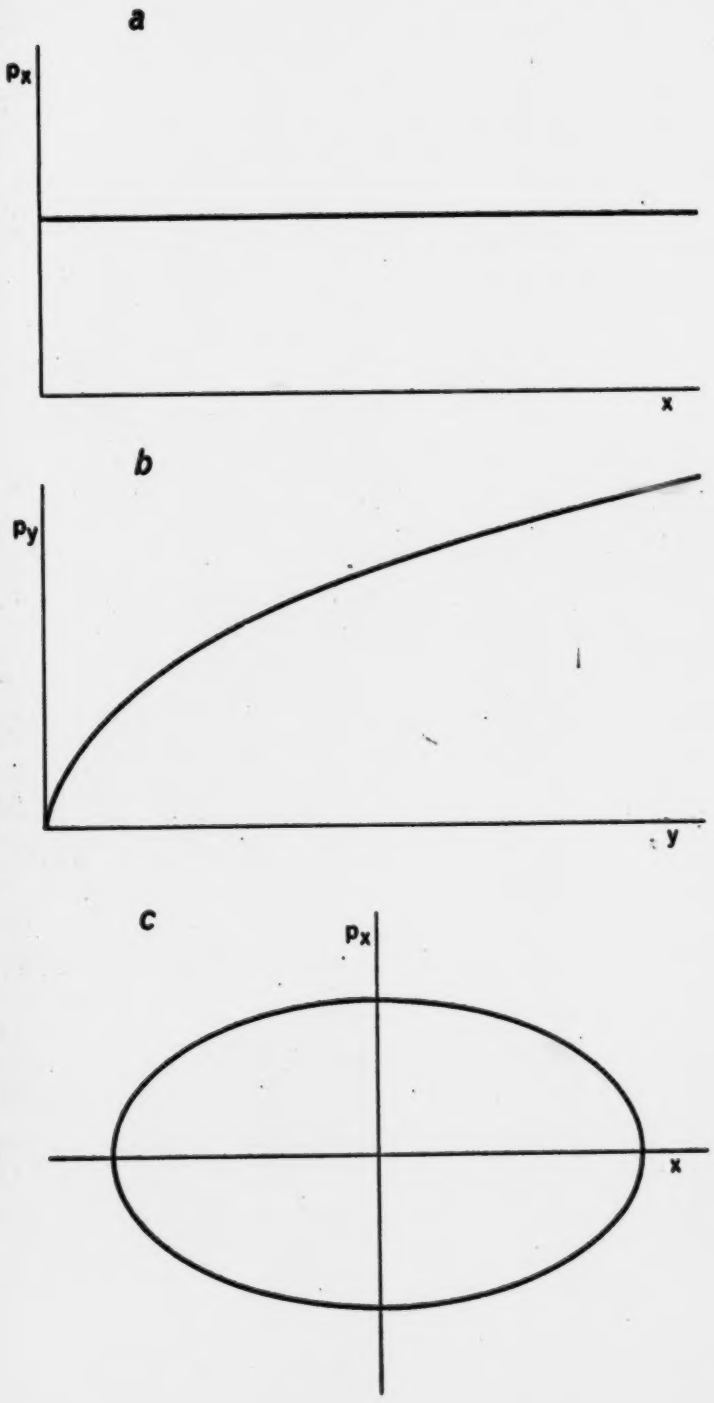


Fig. 4.1 Paths in phase space of some dynamical systems; (a) a free particle; (b) a falling object; (c) a one-dimensional harmonic oscillator.

4.2 MOLECULAR VELOCITIES

By using a combination of physical arguments and the methods of statistical analysis, one can obtain the form of the distribution law for atomic velocities. To make the analysis as simple as possible, we assume that the atoms have no properties other than their (identical) mass, so that a complete description of the state of a single atom can be given by specifying its 3 position coordinates; x, y, z in Cartesian space, and its 3 velocity or momentum coordinates; v_x, v_y, v_z or $p_x = mv_x$, $p_y = mv_y$, $p_z = mv_z$. Although it simplifies the analysis to consider the atoms as if they possessed no internal variables, the resulting distribution of velocities will not depend on this restriction, and it will be possible at a later stage to expand the treatment to include other atomic properties in the statistical description.

In addition to the assumption of point masses, we assume that the gas is isotropic, so that the motions look the same when viewed from any direction. The gas can be isotropic in only a statistical sense, of course, since any particular trajectory has a definite direction in space. But a series of observations of many trajectories is assumed to show no tendency to prefer or avoid certain directions, or to have speeds which are in any way correlated with direction. Since all directions are equivalent, we are able to analyze the three-dimensional motion by studying at first only the component motions parallel to a single direction and later to combine the independent distributions along each direction to make a complete three-dimensional description.

4.3 PHASE SPACE

The dynamical behavior of a point-mass molecule can be described by giv-

ing the history of its 3 position and 3 momentum coordinates. As the 6 variables change with time, the particle can be imagined to move along a path in a six-dimensional hyperspace, called phase space. Certain properties of the paths in phase space, those which are important to the theory of their statistical behavior, can be obtained from their geometry alone, ignoring their behavior as a function of time. In order to become familiar with the concept of phase space, we will consider some examples of motion in one dimension, so that their paths can be drawn as two-dimensional graphs.

A particle moves along the x direction at constant speed. Therefore p_x is constant; x varies. Its graph is shown in Fig. 4.1a. An object falls from rest, with acceleration g. Its momentum $p_y = mgt$, and its displacement $y = \frac{1}{2}gt^2$. Therefore, its phase space trajectory is the parabola $y = p_y^2/2m^2g$, shown in Fig. 4.1b. A mass is attached to a spring, and executes simple harmonic oscillations according to the equation

$$x = x_0 \sin \omega t; \quad p_x = m\omega x_0 \cos \omega t.$$

We can obtain the equation for the corresponding path in phase space by squaring each equation above and using the trigonometric equality.

$$\sin^2 \theta + \cos^2 \theta = 1.$$

Carrying out these operations, we obtain

$$(m\omega x)^2 + p_x^2 = (m\omega x_0)^2.$$

This graph of this equation is an ellipse, shown in Fig. 4.1c.

The importance of the concept of phase space arises from the very powerful technique of analysis of dynamical systems by means of Hamilton's equations. In the Hamiltonian method, coordinates and momenta are in a symmetrical relation with respect to one another; x and p_x, y and p_y, etc. These

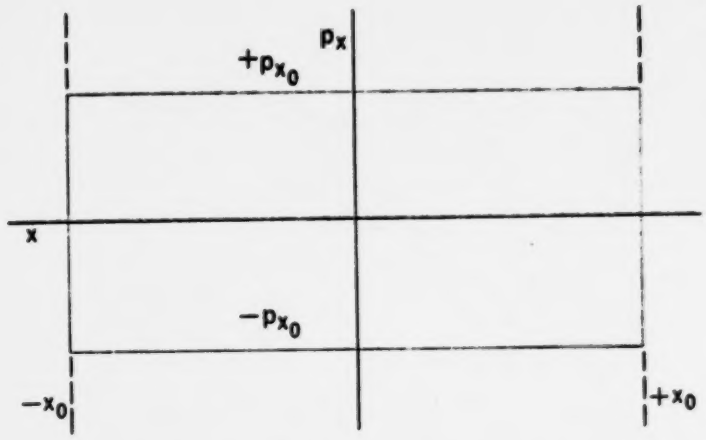


Fig. 4.2 One dimensional motion of a "free" particle confined in a box having perfectly reflecting walls.

pairs of variables are said to be canonically conjugate to each other, and they play symmetric roles in Hamilton's equations of motion.

We are preparing to use the concept of phase space to analyze finite volumes of gases. Therefore, we will modify Fig. 4.1a by confining the particle to a one-dimensional box, so that it only travels between limits, being reflected at the walls. Now, the path is seen in Fig. 4.2: It has become a pair of horizontal lines, one at $+p_x$ as it moves to the right, and one at $-p_x$ as it moves to the left at the same speed. The two horizontal lines are connected by vertical lines at the walls, where the reflections are assumed to take place at precise positions. For real particles and real walls which are slightly compressible, a reflection takes place over a small but nonzero distance, and therefore, the actual trajectory would have slightly rounded corners. In any event, the path is closed.

In the one-dimensional example, the motion of a particle in a box simply repeats in each transit the motion of the previous cycle, but in two- and three-dimensional motion the path is not repeated each time. This is because the total energy is shared among the three components in the different directions, and the component magnitudes will generally vary from one transit to the next. In three dimensions, the par-

ticle moving at constant speed has the total constant energy

$$\epsilon = \frac{1}{2m}(p_x^2 + p_y^2 + p_z^2). \qquad (4.3)$$

As the collisions with the rough walls cause the particle to be reflected at in various directions, the magnitudes of the component momenta p_x, p_y, p_z vary. If all directions are equally probable, there are no restrictions on the individual magnitudes as long as the condition of constant energy is maintained. This corresponds to the range for each component,

$$-\sqrt{2m\epsilon} \le p_i \le +\sqrt{2m\epsilon};\ p_i = p_x,\ p_y,\ p_z.$$

Let us now imagine many identical particles moving in the volume. If the atoms rarely collide with each other, each one will move very nearly as the solitary atom, tracing out an irregular path in the six-dimensional phase space. In the case of many atoms, however, each particle has a very much larger range of momenta than if there were no collisions. In a typical collision between two particles one will be caused to speed up and the other to slow down. The result of a particular collision depends on details of the actual cross sections and directions of approach, but we do not need to consider these details in our statistical analysis.

The momentum available to a single atom now extends up to such a value that the total energy U of the gas is contained in a single particle. The limits of momentum in each dimension are thus

$$-\sqrt{2mU} \le p_x \le \sqrt{2mU}.$$

If there are N identical particles, each one is subject to the same restriction, and each roams over the phase space of larger extent. But they do not do so independently, since all of their energies must add up to U at all times. Therefore, the particles can be represented as N points moving through the six-dimensional phase space, but restricted by the conservation of the total energy to a surface of five dimensions. Projected upon a plane surface in p_x and x, their momentary positions in phase space might resemble the graph in Fig. 4.3. If this drawing were an illustration of an arrangement in coordinate space, it would indicate the variations in density of the molecules, i.e., their distribution in space. But in Fig. 4.3, we see a distribution in phase space, which contains aspects of a distribution in coordinate space combined with a distribution in momentum space.

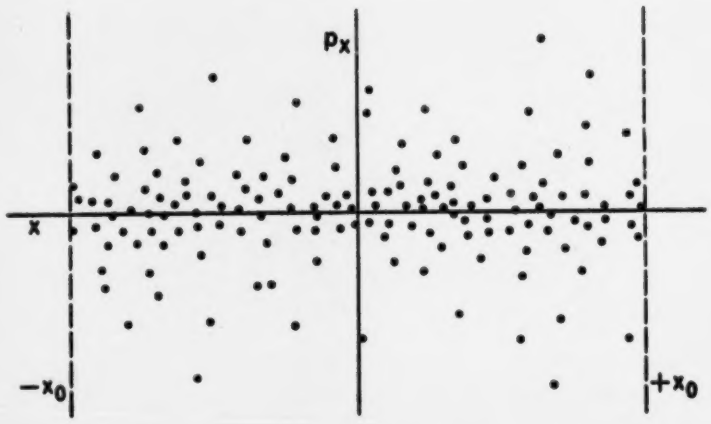


Fig. 4.3 Graph of the instantaneous positions and momenta of a weakly interacting gas in statistical equilibrium.

4.4 DISTRIBUTION IN PHASE SPACE

In order to analyze the distribution in phase space by statistical methods, we will assume that the space is divided into many small cells of equal "volume" $\Delta p_x \cdot \Delta x$. We make the crucial assumption that each cell has an equal probability of occupation - this assumed uniformity of phase space is analogous to the uniformity of coordinate space in the previous section. The condition of uniform probability in phase space, which is the fundamental hypothesis of statistical mechanics, and which is applicable to all identical systems in statistical equilibrium, is discussed in considerable detail in advanced works on statistical mechanics. It rests upon a combination of

physical and mathematical arguments that are beyond the level of this monograph. We shall therefore simply adopt it without any further justification.

We also assume that the probabilities are unchanged by the numbers of atoms already in the cell. This assumption of the independence of probability on cell population is not correct, but is only an approximation which works when the density in phase space is small (in comparison to the reciprocal of Planck's constant h). The approximation is equivalent to the assumption of classical statistical behavior, which is generally valid for gases and other systems at not-too-low temperatures. Under these assumed conditions, if the probability of occupation by a single molecule in any cell is g, the probability of occupation by N molecules is g^N.

For convenience in describing the arrangements of molecules among the cells, we designate the cells by numbering them in sequence, and list their populations thus: n_1 particles in cell 1, n_2 in cell 2, ... n_ν in ν. If the molecules are indistinguishable, the probability for a distinct arrangement is given by a multinomial distribution, as in Chapter 3:

$$p(n_1, n_2, \ldots n_\nu) = w(n_1, n_2, \ldots n_\nu)g^N$$

where

$$w(n_1, n_2, \ldots n_\nu) = \frac{N!}{n_1!\, n_2!, \ldots n_\nu!} \qquad (4.4)$$

is the statistical weight of the arrangement. If N and ν are large numbers the statistical weights of some arrangements are enormously greater than most of the rest. The arrangements of great statistical weight will therefore be those that the gas spends most of its time in. Therefore the gas can be approximately described as being <u>only</u> in its most probable state, ignoring the rest. Fluctuations of the distribution away from the most probable one will occur, analogous to the fluctuations in the ordinary density treated earlier, but we will not consider them here.

Considering only the most probable distribution, we seek to maximize the statistical weight, Eq. (4.4). The distribution is subject to two equations of constraint, however. The first is that the total number of particles is a constant:

$$\sum_{i=1}^{\nu} n_i = N, \qquad (4.5)$$

and the second is that the total energy is constant; if the energy corresponding to the ith cell is ϵ_i,

$$\sum_{i=1}^{\nu} \epsilon_i n_i = U, \text{ where } \epsilon_i = p_i^2/2m. \qquad (4.6)$$

To maximize w, we first express it in more convenient form by taking its logarithm,

$$\ln w = \ln N! - \ln(n_1!n_2! \ldots n_\nu!)$$

$$= \ln N! - \sum_{i=1}^{\nu} \ln n_i!. \qquad (4.7)$$

We use Stirling's approximation for the factorial, Eq. (3.4) (we are assuming that each cell in phase space is large enough so that it contains enough molecules for Stirling's approximation to be accurate). Using the formula for $\ln n!$.

$$\ln n! = (n + \tfrac{1}{2}) \ln n - n + \tfrac{1}{2} \ln 2\pi, \qquad (4.8)$$

substituting in the Eq. (4.7) above,

$$\ln w = (N + \tfrac{1}{2}) \ln N + \frac{(1 - \nu)}{2} \ln 2\pi$$

$$- \sum_{i=1}^{\nu} (n_i + \tfrac{1}{2}) \ln n_i$$

$$= \text{a constant} - \sum_{i=1}^{\nu} (n_i + \tfrac{1}{2}) \ln n_i. \qquad (4.9)$$

For the most probable arrangement, the variation of ln w with respect to small variations of all the populations is zero:

$$d(\ln w) = 0 = -\sum_{i=1}^{\nu} (\ln n_i + 1 + \frac{1}{2n_i}\, dn_i$$

$$\cong -\sum_{i=1}^{\nu} \ln n_i\, dn_i, \quad (4.10)$$

if we neglect terms that are small compared to ln n_i; ln $n_i \gg 1$.

The small variations in the cell populations are constrained to be consistent with the constant N and U of the entire collection of particles, by the method of undetermined Lagrance multipliers. This is a method which can be used when a function of several variables has either a maximum or a minimum, and where there are additional equations of condition linking the variables. Each independent equation among the variables reduces by 1 the number of independent degrees of "freedom" of the function. The method of Lagrange multipliers is a simple technique for bringing the equations of constraint into a single functional equation. Its application will be seen in our case: By differentiating Eq. (4.5) and Eq. (4.6) with respect to the populations in each cell,

$$\sum_{i=1}^{\nu} dn_i = 0, \text{ and } \sum_{i=1}^{\nu} \epsilon_i dn_i = 0,$$

we obtain two equations linking the variations among the n_i. They can be incorporated into Eq. (4.10) by multiplying each by an undetermined parameter and adding all three together;

$$\alpha' \sum_{i=1}^{\nu} dn_i = 0$$

$$+ \beta \sum_{i=1}^{\nu} \epsilon_i\, dn_i = 0$$

$$+ \sum_{i=1}^{\nu} \ln n_i\, dn_i = 0;$$

$$\sum_{i=1}^{\nu} (\ln n_i + \alpha' + \beta\epsilon_i) dn_i = 0.$$

Since this equation does not depend on the choice of cell size or the numbers of particles in each cell, the term in parentheses must vanish for each cell independently:

$$(\ln n_i + \alpha + \beta\epsilon_i) = 0.$$

Taking antilogs

$$n_i = \alpha \exp -\beta\epsilon_i, \quad (4.11)$$

where $\alpha = e^{-\alpha'}$

Equation (4.11) is one of the most famous equations in physics. It was obtained by Boltzmann (1844-1906), in 1868. The term exp $-\beta\epsilon_i$ is known as the Boltzmann factor, and it has applicability to a wide variety of systems in statistical equilibrium, not only the point-mass atoms of a gas. For, if we review the analysis leading up to this point, we find that it depends on only the following conditions and assumptions:

(a) There are many identical particles in statistical equilibrium.

(b) All volumes of phase space are equally probable.

(c) The probability of occupation of a cell is independent of the number of particles in the cell. (Note that the cell size is arbitrary in our discussion.)

(d) The populations in each cell are large enough for Stirling's approximation to be used.

These conditions are also satisfied by many other systems, including the vibrations of atoms in a crystal, the states of atomic moments in a magnetic solid, and the conduction electrons in a semiconductor. In the present monograph, however, we are restricting our attention to the case of a gas, which will now be examined in detail.

4.5 THE MAXWELL-BOLTZMANN GAS

Equation (4.11) seems a little artificial for the gas of atoms, since it describes the distribution in terms of the populations of discrete cells in

phase space. These cells are only an artifact we introduced to make the various regions of the total phase volume distinguishable by subscripts. However, we now see that the subscripts are unnecessary, since the energy ϵ is all that the populations depend on, through the Boltzmann factor itself. Although the phase space coordinates of the cell are not important, the cell volume is: By condition (b) above, the population of a cell is directly proportional to its volume. Therefore, the number of atoms in a certain region of phase space is proportional to the volume of the region and to its Boltzmann factor. If we define the population of a cell of energy ϵ and volume $\Delta p_x \Delta p_y \Delta p_z \Delta x \Delta y \Delta z$ to be Δn, Eq. (4.11) is transformed to

$$\Delta n(\epsilon, \Delta p_x \Delta p_y \ldots \Delta z) = \sigma e^{-\beta\epsilon} \Delta p_x \Delta p_y \Delta p_z \Delta x \Delta y \Delta z. \quad (4.12)$$

For infinitesimally small volumes, we have the differential form

$$dn(\epsilon, dp_x dp_y dp_z dV) = \sigma e^{-\beta\epsilon} dp_x dp_y dp_z dV, \quad (4.13)$$

where $dV = dxdydz$ is an infinitesimal volume in coordinate space.

Now, we can apply Eq. (4.13) directly to the case of the point-mass gas. Since the atoms are assumed to move in a field-free region, the energy is not a function of the position of the particle. Therefore, the populations are also independent of their locations in the container, and we can combine them by integrating over the entire volume V:

$$dn(\epsilon, dp_x dp_y dp_z) = \sigma V e^{-\beta\epsilon} dp_x dp_y dp_z. \quad (4.14)$$

Since the particles are assumed to collide infrequently, each atom moves essentially as an independent particle. Therefore, the energy is related to the momentum by the free particle expression, Eq. (4.3). Substituting for ϵ, Eq. (4.14) becomes

$$dn(\epsilon, dp_x dp_y dp_z) =$$

$$\sigma V \exp\left[-\beta\left(\frac{p_x^2}{2m} + \frac{p_y^2}{2m} + \frac{p_z^2}{2m}\right)\right] dp_x dp_y dp_z =$$

$$\sigma V \left[\exp\left(-\frac{\beta p_x^2}{2m}\right) dp_x\right]\left[\exp\left(\frac{-\beta p_y^2}{2m}\right) dp_y\right]$$

$$\times\left[\exp\left(\frac{-\beta p_z^2}{2m}\right) dp_z\right]. \quad (4.15)$$

This result contains the familiar form of the Gaussian distribution: We see that it is the product of three Gaussian distributions, one in each momentum variable. By comparing the momentum distributions with the general expression, Eq. (3.17), we can obtain the average momentum $\langle p_x \rangle$ and the average squared momentum $\langle p_x^2 \rangle$ by inspection. Comparison of the exponents shows that the term which would correspond to μ in the Eq. (3.17) is absent in each factor of the momentum distribution. Since μ is equal to the average of the variable itself, this means that the average of each momentum component vanishes:

$$\langle p_x \rangle = \langle p_y \rangle = \langle p_z \rangle = 0. \quad (4.16)$$

It is a reassuring result, since the gas was originally presumed to be in equilibrium with its stationary container. Comparison of the remaining terms of the exponents yields

$$\frac{\beta p_x^2}{2m} = \frac{p_x^2}{2\sigma_x^2}; \quad \frac{\beta p_y^2}{2m} = \frac{p_y^2}{2\sigma_y^2}; \quad \frac{\beta p_z^2}{2m} = \frac{p_z^2}{2\sigma_z^2}.$$

Therefore, $\sigma_x^2 = \sigma_y^2 = \sigma_z^2 = \sigma^2 = m/\beta$. (4.17)

Thus, the dispersion in the momentum is the same in each direction. Since the average momentum in each direction is zero, we have a simple relation between the dispersion and the average of the squared components. By Eq. (3.12) we see that in this case

$$\sigma^2 = \langle p_x^2 \rangle = \langle p_y^2 \rangle = \langle p_z^2 \rangle = m/\beta. \quad (4.18)$$

Therefore, the averages are simply related to the Lagrange multiplier β and the atomic mass. The parameter β, which has been undetermined so far, can now be related to physical properties by comparison with experiment. For we found in Chapter 1 that Charles' law implies a simple proportionality between the kinetic energy and the temperature T on the absolute scale. That connection leads us to the discovery of β: by Eq. (1.12) and Eq. (4.18), we have

$$\langle \epsilon \rangle = \frac{p_x^2}{2m} + \frac{p_y^2}{2m} + \frac{p_z^2}{2m} = \frac{3}{2\beta} = \frac{3}{2} kT.$$

And therefore, we obtain the gloriously simple and profound relation,

$$\beta = 1/kT. \qquad (4.19)$$

Before turning to a discussion of the significance of the result, we can effect a definition of the other Lagrange multiplier by a similar comparison with the general form of the Gaussian distribution. By this comparison it is found that

$$\alpha = \frac{N}{V(2\pi mkT)^{3/2}}. \qquad (4.20)$$

Therefore, the basic properties of the statistical distribution of momentum have been found, and we can write the distribution law in terms of identified quantities:

$$dn = \frac{N}{(2\pi mkT)^{3/2}} \exp\left[-\frac{p^2}{2mkT}\right] dp_x dp_y dp_z \qquad (4.21)$$

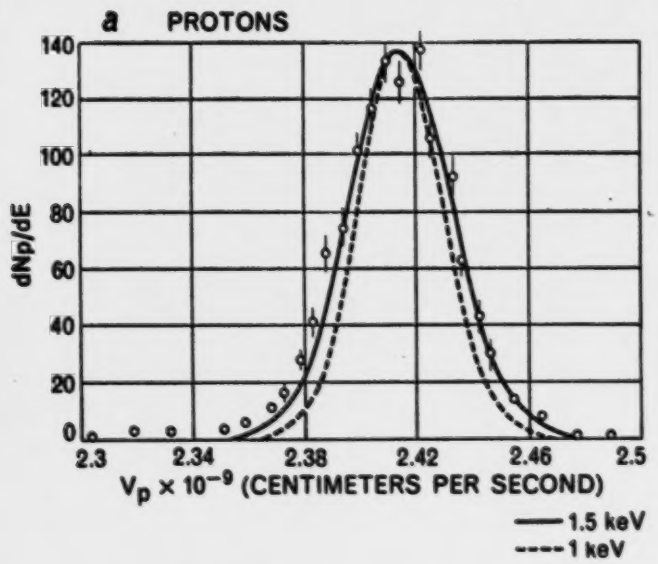
where we have written p for the total momentum $\sqrt{p_x^2 + p_y^2 + p_z^2}$.

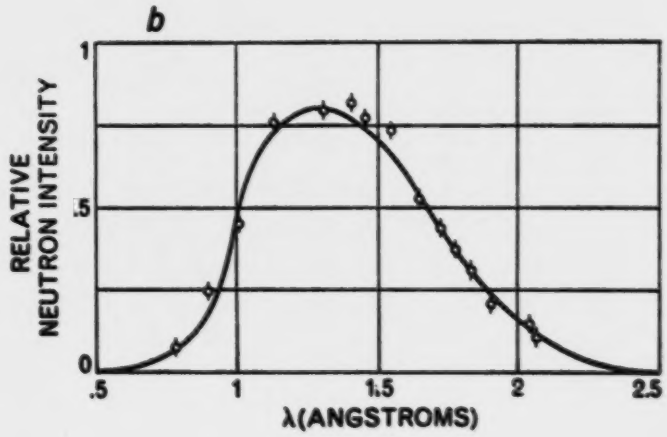
By a combination of reference to experiments on gross volumes of gases, a belief that gases are composed of atoms, and a statistical analysis of their random motions, we have arrived at a formula for the velocities of heat motion of the atoms in a gas.

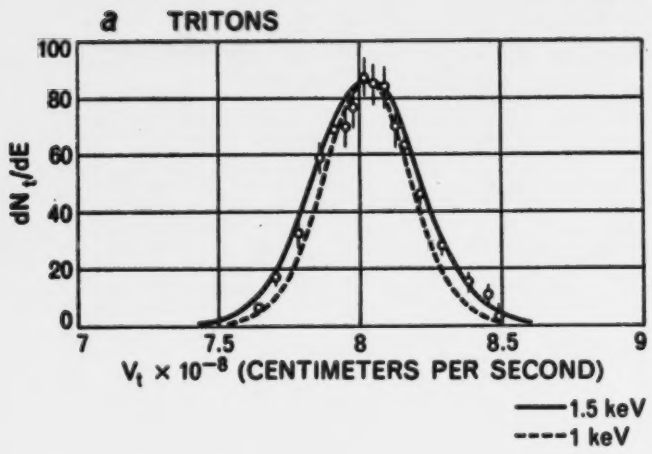
Equation (4.21) was originally derived in 1856 by Clerk Maxwell (1831-1879), whose analysis followed very different lines. Our derivation is patterned after Boltzmann's. In either the form above or in one of several related forms, it is called the Maxwell-Boltzmann distribution law. It was confirmed experimentally by several researchers in this century: One of the first was Otto Stern, in 1920. Since then, the Maxwell-Boltzmann distribution has been tested for other "gases," including plasmas and neutrons in a reactor. Some experimental comparisons are shown in Fig. 4.4.

According to the Maxwell-Boltzmann distribution law, the momentum distribution has a width that is controlled by the temperature. If more heat energy is given to the atoms, they will, by means of their collisions, share it among themselves according to statistical laws within a certain relaxation time. The added heat raises the average energy of all the atoms, and although the average momentum in each direction is zero, the average squared momentum increases. This increases the dispersion of the distribution, so that the variations in the momenta of atoms picked at random are greater at higher temperatures. We can liken these changes with temperature to the variations that would be seen in an error curve, if the measuring instrument had a variable sensitivity. The coarseness of the "measurement" is an effect of the many collisions which make up the statistical equilibrium: As the average energy increases, so does the dispersion. Cooling the gas improves the precision of the "measurement," but it also reduces the quantity itself. As the temperature is reduced, the curve becomes sharper, and if the Maxwell-Boltzmann distribution were to continue to be obeyed at all temperatures, it implies that one could measure precisely $\langle \epsilon \rangle = 0$ at $T = 0$. However, long before then, real gases liquefy, and the assumptions leading to the distribution law become poor approximations of the real situation. Other forms of heat motion take place in liquids and solids, and they may be analyzed by sta-

tistical analyses based on the same approximations we have used here. On such a basis, the heat motions in liquids and solids also vanish at T = 0. But it was realized early in the twentieth century that several of the assumptions of the classical theory are not strictly correct. In the form in which they are given following Eq. (4.11), they are reasonable approximations for relatively high temperatures and low densities, but for dense and cold materials, the motion must be analyzed in terms of quantum mechanical laws. One important effect of quantum mechanical behavior can be







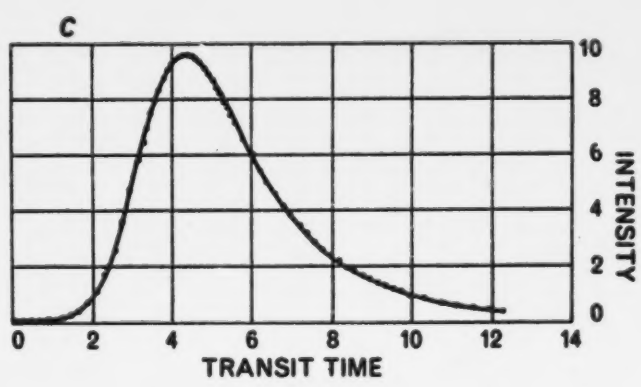


Fig. 4.4 Measured distributions of various properties of some systems of weakly interacting particles, compared with theoretical curves derived from the Maxwell-Boltzmann law.
(a) Velocity spectra of protons (a.1) and tritons (a.2) from the d-d reaction in the Scylla experiment on fast magnetic compression of a plasma, obtained by D. E. Nagle, W. E. Quinn, W. B. Reisenfeld, and W. Leland, Phys. Rev. Letters 3, 318 (1959). The curves are drawn according to theoretical distributions corresponding to temperatures $T = 11.6 \times 10^6$ °K (1.0 keV) and 17.4×10^6 °K (1.5 keV).
(b) Intensity distribution of neutrons emitted in thermal beam of a reactor, as a function of neutron wavelength $\lambda(\lambda = h/mv)$, measured by J. G. Dash and H. S. Sommers, Jr., Rev. Sci. Instr. 24, 91 (1953). The curve is calculated for a beam effusing from a Maxwellian source at 320 °K, modified by a spectrometer transmission function.
(c) Measured transmission curve and calculated Maxwell-Boltzmann transmission curve for a beam of potassium atoms effusing from an oven at 157 °C. Abscissa is approximately equal to transit time of atoms through velocity selector. Measurements were made by P. M. Marcus and J. H. McFee, Recent Research in Molecular Beams, I. Estermann (ed.), (Academic Press, New York, 1959).

stated in closing: The motions at T = 0 are small, but not zero. As absolute zero is approached, the atoms slow down to some minimal momenta and energies. When all the heat energy is removed, a minimum and unextractable amount remains. This so-called zero-point motion implies a residual dispersion in the momentum, a direct consequence of the laws of quantum mechanics.

EPILOGUE

In this monograph we have attempted to show the march of scientific progress in the field of kinetic theory of gases. Experiment confirmed theory brilliantly: perhaps the student might have expected it. Textbooks and monographs tend to have this sort of happy ending, a neatly packaged rationalization or thesis. Economy of the teacher's and student's time usually demands that all of the mistakes and fruitless paths in the history of the science must must be ignored. This lends an unreal and distorted complexion to one's impressions of the subject and to scientific achievement in general. The development of our ideas about heat motions of atoms is a case in point. From the first insights of the Greek atomists to the experimental confirmation of the detailed statistical theory took more than 2000 years. During much of this time men were attempting in one way or another to rationalize the universe in general and the nature of fire in particular. Most of them were wrong, or at least not part of the straight line of progress sketched in our description. And yet, the distance we traveled in two millennia is not so enormous that there were no glimmerings of modern ideas at the very beginning. Here is what Lucretius wrote in the first century B.C.:

> These atoms, which are separated from each other in the infinite void and distinguished from each other in shape, size, position and arrangement, move in the void, overtake each other and collide.

And also

> Observe what happens when sunbeams are admitted into a building and shed light in shadowy places. You will see a multitude of tiny particles mingling in a multitude of ways in the empty space within the light of the beam, as though continuing in everlasting conflict, rushing into battle rank upon rank with never a moment's pause in a rapid sequence of unions and disunions. From this you may picture what it is for the atoms to be perpetually tossed about in the void.

BIBLIOGRAPHY AND SUGGESTED READINGS

HISTORY

S. Sambursky, The Physical World of The Greeks, (translated by Merton Dagut), (Routledge and Kegan Paul, Ltd., London, 1956).

Benjamin Farrington, Greek Science (Penguin Books, Baltimore, Maryland, 1953).

H. T. Pledge, Science Since 1500, Harper Torchbooks, Harper and Row, Publishers, New York, 1959).

Lancelot Hogben, Statistical Theory (Allen and Unwin, London, 1957).

MATHEMATICS OF PROBABILITY

George P. Wadsworth and Joseph G. Bryan, Introduction to Probability and Random Variables (McGraw-Hill Book Company, New York, 1960).

Lyman G. Parratt, Probability and Experimental Errors in Science (John Wiley & Sons, Inc., New York, 1961).

PHYSICS

E. H. Kennard, Kinetic Theory of Gases (McGraw-Hill Book Company, New York, 1938).

R. D. Present, Kinetic Theory of Gases, (McGraw-Hill Book Company, New York, 1958).

M. W. Zemansky, Heat and Thermodynamics, (McGraw-Hill Book Company, New York, 1957) 4th ed.

P. M. Morse, Thermal Physics, (W. A. Benjamin, Inc., New York, 1964).